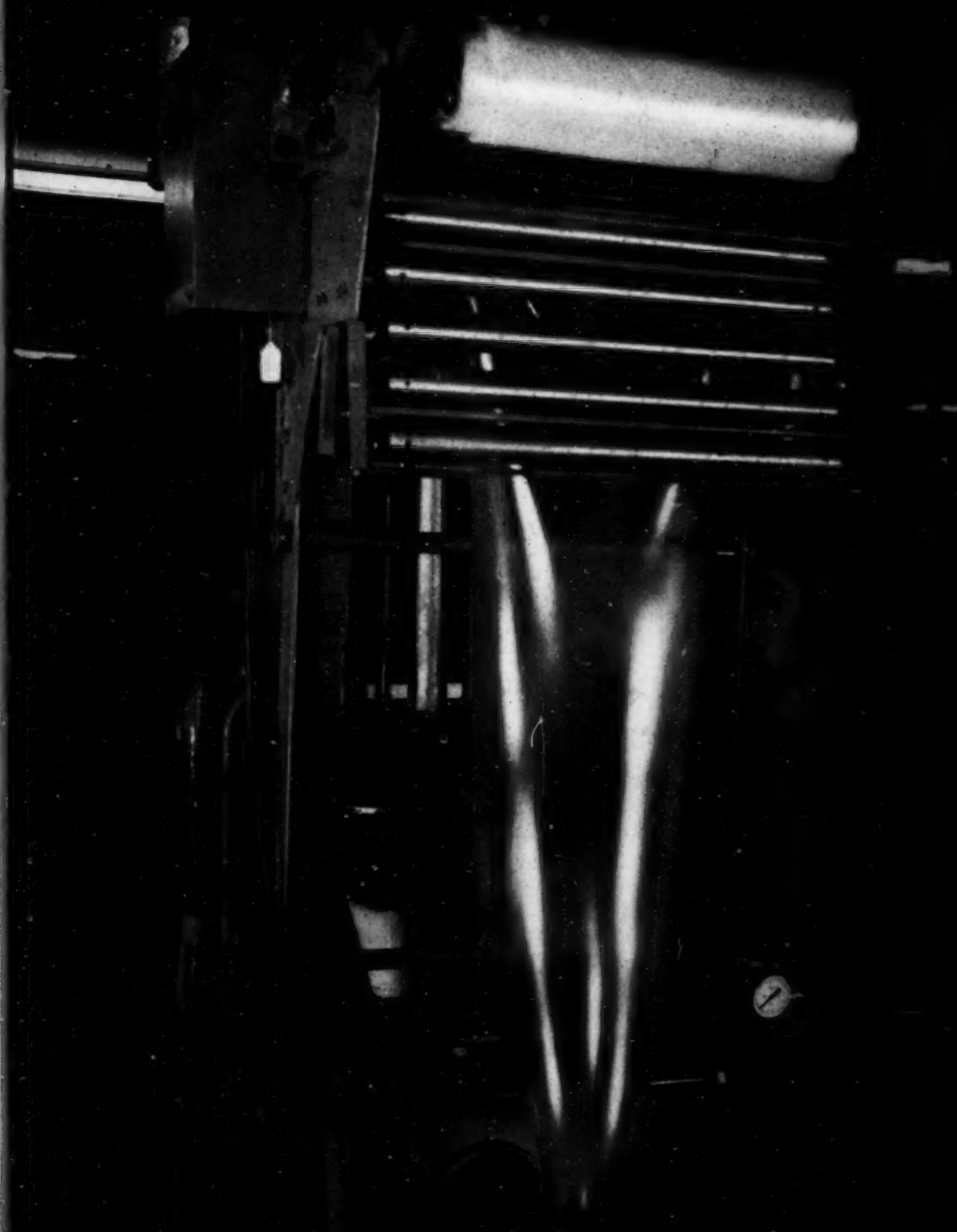
JOURNAL OF APPLIED PHYSICS



JOURNAL OF APPLIED PHYSICS

Vol. 25, No. 9 SEPTEMBER, 1954

In This Issue

Rheology Symposium **Page**

Composition of Apparent Shearing Forces During Shear Degradation of Polymers A. B. Bestul **1069**

Relaxation Time Spectrum of Dough and the Influence of Temperature, Rest, and Water Content J. R. Cunningham and I. Hlynka **1075**

Constant Stress Elongation of Soft Polymers: Plasticizer Studies Marshall R. Hatfield and George B. Rathmann **1082**

Some Rheological Properties of Concentrated Polyisobutylene Solutions F. J. Padden and T. W. DeWitt **1086**

Elastoviscous Properties of Amorphous Polymers in the Transition Region. II. Stress-Relaxation of Butadiene-Styrene Copolymers E. Catsiff and A. V. Tobolsky **1092**

Rheological Unit in Raw Elastomers M. Mooney and W. E. Wolstenholme **1098**

Further Dynamic Investigations on Polymers Wladimer Philippoff **1102**

Stress and Strain (Discussion of Dr. Philippoff's Paper) W. F. G. Swann **1108**

Effect of Rheological Behavior on Thermal Stresses Alfred M. Freudenthal **1110**

Flow of Melts in "Crosshead"-Slit Dies; Criteria for Die Design James F. Carley **1118**

Rotational Plastometry Applied to Molten Polyethylene F. D. Dexter **1124**

Internal Friction and Dynamic Modulus of Cold-Worked Metals A. S. Nowick **1129**

Rheology of Suspensions Keith H. Sweeny and Richard D. Geckler **1135**

Regular Contributed Original Research

Analysis of Systems Involving Difference-Differential Equations T. M. Burford **1145**

Transient Response of a *p-n* Junction Benjamin Lax and S. F. Neustadter **1148**

Lower Modes of a Concentric Line Having a Helical Inner Conductor Louis Stark **1155**

Minimum Noise Figure of Traveling-Wave Tubes with Uniform Helices J. R. Pierce and W. E. Danielson **1163**

Dielectric Constant and Loss Measurements on Barium Titanate Single Crystals While Traversing the Hysteresis Loop M. E. Drougard, H. L. Funk, and D. R. Young **1166**

Solute Redistribution by Recrystallization Robert G. Pohl **1170**

Matrix Analysis of Linear Time-Varying Circuits Louis A. Pipes **1179**

Statistical Properties of the Output of a Frequency Sensitive Device George R. Arthur **1185**

Amplification Factor and Perveance of an Elliptic Triode S. Deb and G. S. Sanyal **1196**

Letters to the Editor:

Formation of Single Crystal Silicon Fibers Everett R. Johnson and James A. Amick **1204**

Laminar Flow of an Electrically Conducting Incompressible Fluid in a Circular Pipe Shih I. Pai **1205**

Ultimate Vacuum in a Vacuum-Enclosed Ionization Gauge L. J. Varnerin, Jr., and D. White **1207**

Note on the Propagation of Explosion-Produced Air Shocks M. Birk, Y. Manheimer, and G. Nahmani **1208**

Errata: Velocity Attenuation of Explosive-Produced Air Shocks Jacob Savitt and R. H. F. Stresau **1208**

Index to Advertisers **xx**

ROBERT L. SPROULL
Editor

ELMER HUTCHISSON,
EARLE C. GREGG,
Retiring Editors

Board of Editors
MAURICE A. BIOT
RAY E. BOLZ
WALLACE J. ECKERT
ROBERT D. HEIDENREICH
LOUIS A. PIPES
ROMAN SMOLUCHOWSKI
W. R. SMYTHE
ARTHUR V. TOBOLSKY
ALVIN M. WEINBERG

Publication of the
American Institute of Physics

OFFICERS
FREDERICK SEITZ,
Chairman
GEORGE B. PEGRAM,
Treasurer

GOVERNING BOARD
ROBERT F. BACHER
J. W. BUCHTA
KARL K. DARROW
S. A. GOUDSMIT
DEANE B. JUDD
HUGH S. KNOWLES
C. C. LAURITSEN
W. H. MARKWOOD, JR.
WILLIAM F. MEGGERS
PHILIP M. MORSE
BRIAN O'BRIEN
HARRY F. OLSON
GEORGE B. PEGRAM
ERIC RODGERS
RALPH A. SAWYER
FREDERICK SEITZ
WILLIAM SHOCKLEY
RICHARD M. SUTTON
J. H. VAN VLECK
MARK W. ZEMANSKY

Consultant on Printing
MELVIN LOOS

ADMINISTRATIVE STAFF
HENRY A. BARTON,
Director
WALLACE WATERFALL,
Executive Secretary
THEODORE VORBURGER,
Advertising Manager
RUTH F. BRYANS,
Publication Manager
EDITH I. NEFTEL
Circulation Manager
KATHRYN SETZE
Assistant Treasurer

The JOURNAL OF APPLIED PHYSICS, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as PHYSICS. The JOURNAL OF APPLIED PHYSICS publishes editorials and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Robert L. Sproull, Editor, Rockefeller Hall, Cornell University, Ithaca, New York. Unless otherwise stated, submission of a manuscript is a representation that it has been neither copyrighted, published, nor submitted for publication elsewhere. The authors' institutions are requested to pay a **publication charge** of $8 per page which, if honored, entitles them to 100 reprints without further charge. Instructions will be sent with galley proofs.

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price	U.S. and Canada	Elsewhere
To members of the American Institute of Physics and Affiliated Societies	$10.00	$11.00
To all others	12.00	13.00

Back Number Prices

Yearly back number rate when complete year is available: $14.00. Single copies: $1.25 each.

Changes of address should be sent to the American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics. **Advertising rates** supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized May 2, 1932.

Journal
of
Applied Physics

Volume 25, Number 9 September, 1954

Rheology Symposium

The papers in the first half of this issue were presented at the Annual Fall Meeting of the Society of Rheology. This meeting was held at the Hotel New Yorker in New York City, October 29, 30, and 31, 1953.

Composition of Apparent Shearing Forces During Shear Degradation of Polymers*†‡

A. B. Bestul
National Bureau of Standards, Washington, D. C.
(Received April 12, 1954)

An investigation has been made of the degradation of Vistanex polyisobutenes having viscosity-average molecular weights from about 500 000 to 2 500 000 by shearing in solution in cetane at polymer concentrations from 5 to 20 weight percent, at nominal rates of shear from 9000 to 66 000 sec^{-1}, and at temperatures from 20° to 80°C. The initial apparent shearing loads, observed at the beginning of the degradation process, vary in unexpected ways with variations in nominal rate of shear, temperature, concentration, and initial molecular weight of the polymer. In some cases the observed variations are opposite to those which would be expected on the basis of viscoelastic resistance to shearing. These results can be explained on the basis of the hypothesis that the observed apparent shearing forces include some contribution associated with the energy used up in breaking chemical bonds in the polymer molecules during the degradation process. If this hypothesis is correct, valid viscosity values cannot be obtained directly from observed apparent shearing forces in viscometry measurements under conditions where shear degradation occurs.

1. INTRODUCTION

THIS paper reports findings which indicate that apparent shearing forces in viscometric measurements on high polymer systems under conditions where shear degradation occurs may include a contribution which is not associated with overcoming viscoelastic resistance. These findings are reported here because of their importance to the valid utilization of viscometry measurements under these conditions.

There has recently been carried out in this laboratory an extensive program of measurements on the mechanical degradation by capillary shearing of polyisobutenes in concentrated solutions in cetane. These measurements were made as part of a study directed at elucidating the kinetics and the energy requirements of this type of degradation, with a view to developing a method for obtaining some sort of quantitative characterization of the extent and strength of intermolecular entanglements in linear amorphous high polymer systems. The degradation was investigated as a function of concentration in the range of 5 to 20 weight percent, of nominal rate of shear in the range of 9000 to 66 000 sec^{-1}, of temperature in the range of 20 to 80°, and of initial viscosity-average molecular weight in the range of 490 000 to 2 520 000. Every condition for each of these variables was not studied over the full range stated for each of the other three variables, however.

The degradation was accomplished by repeatedly forcing a solution through a capillary tube at a fixed rate of flow. The force required to produce the fixed rate of flow was measured every other time the solution

* Extracted from a thesis submitted to the Graduate School of the University of Maryland in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

† This work was performed as part of a research project sponsored by the Reconstruction Finance Corporation, Office of Synthetic Rubber, in connection with the Government Synthetic Rubber Program.

‡ Presented before the 1953 annual meeting of the Society of Rheology in New York City, on October 29–31, 1953.

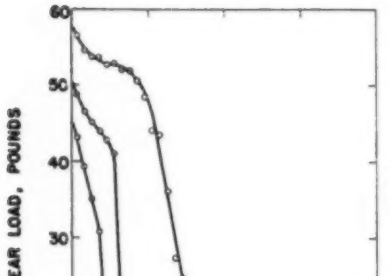


FIG. 1. Shear load decrease for 10 percent *B*-100 Vistanex in cetane at 40° and various values of *D*.

was forced through the capillary, thus providing a measure of the apparent shearing force as a function of the amount of shearing to which a sample of solution had been subjected. In these measurements the initial apparent shearing force, during the first time a solution was forced through the capillary, depended in unexpected ways upon some of the above-mentioned variables. In some cases the variation of the initial apparent shearing force was opposite to that which would be expected if this force originated entirely from viscoelastic resistance to flow. In addition, it was determined from these measurements, considered together with complementary intrinsic viscosity measurements, that the course of the apparent shearing force as a function of the duration of shearing failed to give a true indication of the course of the shear degradation process.

The explanation assigned for this failure is that the apparent shearing forces include a contribution associated with energy used up in the degradation process, as well as a contribution associated with overcoming viscoelastic resistance. Therefore, the apparent shearing forces observed during the degradation process are not considered to be proportional to any kind of viscosity as it is usually conceived, since they include the additional contribution of a quite different nature. This explanation is also in accord with the failure of the initial apparent shearing forces to behave in the manner expected of viscoelastic forces, since they are considered to include forces other than viscoelastic ones.

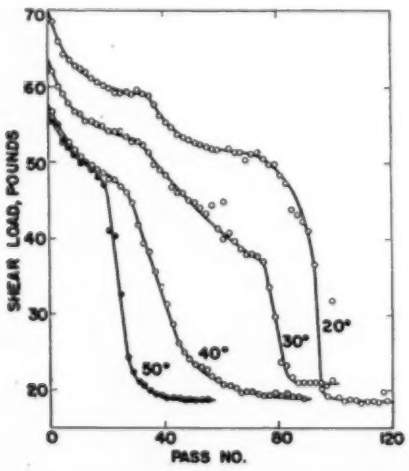


FIG. 2. Shear load decrease for 10 percent Vistanex *B*-140 in cetane at $D=66\,000$ sec^{-1} and various temperatures.

In order to utilize viscometry measurements validly, it is universally necessary to know the composition of the observed apparent shearing forces. As a demonstration of what these forces may include in the presence of shear degradation, this separate paper presents the results from the above-mentioned measurements on the behavior of the apparent shearing forces with variations in the four variables mentioned; although other parts of the investigation are reported elsewhere.[1] No reasons are seen why the principles of the findings reported here should not apply in other instances of viscometry measurements on high polymer systems in the presence of shear degradation, and this paper discusses some possible consequences of the presence of the additional contribution to the apparent shearing forces.

TABLE I. Polyisobutenes investigated.

Vistanex No.	$[\eta]$, in cetane at 30°, dl/g	$\bar{M}_v \times 10^{-6}$
B-40	1.23	0.49
B-60	1.85	0.91
B-80	2.00	1.02
B-100	2.88	1.75
B-120	3.04	1.90
B-140	3.70	2.52

2. EXPERIMENTAL

The polymers investigated are six unfractionated commercial Vistanex polyisobutenes which are listed with their designations, their intrinsic viscosities in cetane at 30°, and the corresponding viscosity-average molecular weights in Table I. These polymers were supplied by the Enjay Company, Inc. The polymers were investigated in cetane (*n*-hexadecane) solution.

Ten percent solutions of the six polymers, and 5, 15, and 20 percent solutions as indicated later of some of them, were prepared in cetane by placing weighed quantities of polymer and of solvent in cylindrical green glass jars and rolling the jars at about six revolutions per minute until the components appeared homogeneous on visual inspection. The five and the ten percent mixtures achieved complete solution in about ten days of rolling. It was necessary to prepare the 15 percent solutions by rolling at 60° in order to achieve reproducible flow properties for them within a reasonable time. At this temperature the components achieved

[1] A. B. Bestul (to be published).

complete solution in about two weeks of rolling. Several ten percent solutions were prepared with the addition of 0.5 percent phenyl-beta-naphthylamine as a free radical acceptor, several were prepared with dried cetane, and several with moisture saturated cetane. None of these last-mentioned variations had any effect on the results of the shear degradation measurements.

Degradation runs were made with the McKee consistometer, which is a capillary viscometer with controlled flow rate and which has already been described in the literature.[2,3] Intrinsic viscosities were determined with an Ubbelohde suspended level viscometer[4] in which the reservoir had been enlarged so that it could be used as a mixing chamber.

Intrinsic viscosities of polyisobutene in cetane at 30° were converted to viscosity-average molecular weights by Flory's theoretical relation[5] using parameters as evaluated from his measurements on intrinsic viscosity-

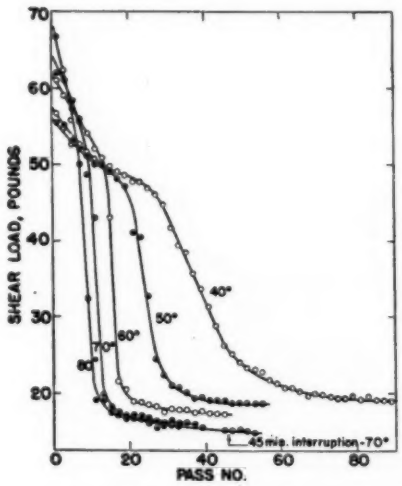


FIG. 3. Shear load decrease for 10 percent Vistanex B-140 in cetane at $D=66\,000$ sec^{-1} and various temperatures.

temperature relationships for polyisobutene in various solvents including cetane at temperatures from 0 to 100°.[6]

Degradation runs were made on the five following sets of solutions under the shearing conditions stated:

(1) 10 percent B-100 at 40° and 66 000, 33 000, 19 000, and 9000 sec^{-1} nominal rate of shear;

(2) 10 percent B-140 at 66 000 sec^{-1} and 20, 30, 40, 50, 60, 70, and 80°;

(3) 5, 10, 15, and 20 percent B-100 at 40° and 66 000 sec^{-1};

[2] S. A. McKee and H. S. White, A. S. T. M. Bull. **No. 153**, 90 (1948).
[3] S. A. McKee and H. S. White, J. Research Natl. Bur. Standards **46**, 18 (1951).
[4] J. Ubbelohde, J. Inst. Petroleum Technol. **19**, 376 (1933).
[5] P. J. Flory and T. G. Fox, Jr., J. Am. Chem. Soc. **73**, 1904 (1951).
[6] T. G. Fox, Jr. and P. J. Flory, J. Am. Chem. Soc. **73**, 1909 (1951).

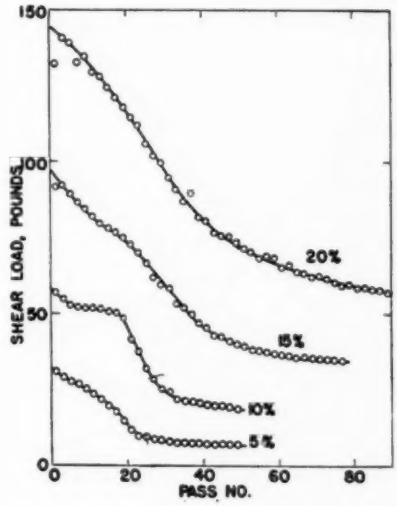


FIG. 4. Shear load decrease for solutions of B-100 Vistanex in cetane at 40°, $D=66\,000$ sec^{-1}, and various concentrations.

(4) 10 percent B-40, B-60, B-80, B-100, B-120, and B-140 at 40° and 66 000 sec^{-1};

(5) and 15 percent B-80, B-100, and B-140 at 40° and 66 000 sec^{-1}.

In a degradation run a solution was repeatedly forced back and forth through the shearing capillary of the McKee consistometer and the driving load proportional to the pressure drop through the capillary was recorded for every other successive passage of the solution through the capillary. A decrease of shearing load with successive passes of the solution through the shearing capillary was taken as an indication of the occurrence of degradation. The behavior of the shearing load with increasing number of passes through the capillary is

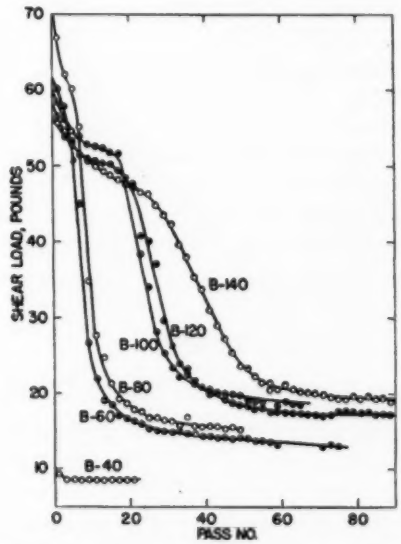


FIG. 5. Shear load decrease for 10 percent solutions of various molecular weight Vistanexes in cetane at 40° and $D=66\,000$ sec^{-1}.

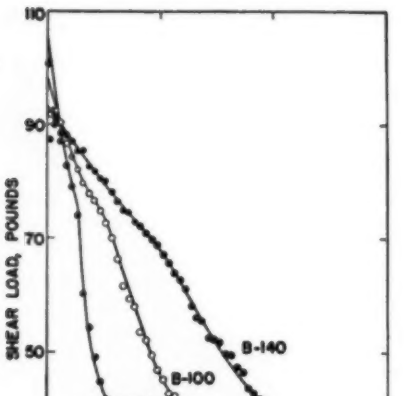


FIG. 6. Shear load decrease for 15 percent solutions of various molecular weight Vistanexes in cetane at 40° and 66 000 sec^{-1}.

shown in Figs. 1 through 6 for the respective conditions previously stated.

The curves in Figs. 1 through 6 are not thought to represent the course of the degradation process. Therefore, intrinsic viscosity determinations were resorted to in order to obtain such a representation for the degradation of ten percent *B*-100 at 40° and 66 000 sec^{-1} as an example of the general nature of the course of the process. Intrinsic viscosity determinations were made on the undegraded polymer and on the polymer at eleven different stages of the degradation process. For determinations during the course of degradation, the shearing was stopped at the desired stage and the solution was removed from the consistometer. The solution was then successively diluted in the chamber of the Ubbelohde viscometer by addition of appropriate amounts of cetane to concentrations of 0.5, 0.25, 0.125, and 0.0625 weight percent, and the viscosity of each of the solutions so obtained was measured at 30° with the viscometer. Viscosity-average molecular weights calculated from these intrinsic viscosities are plotted *versus* the number of times the solution had been forced through the shearing capillary in Fig. 7, where they are represented by the filled circles. For comparison, the unfilled circles in Fig. 7 represent the shear loads during the course of degradation under these same conditions.

3. RESULTS

3.1 Shear Load Decrease During Degradation

The general nature of the plots of shear load *versus* number of passes of a solution through the capillary, as shown in Figs. 1 through 7, is that the shear load decreases comparatively slowly with increasing number of passes during the earlier passes, and this decrease becomes very much more rapid at some later stage of the degradation process, after which the shear load levels off to some essentially constant value. If these curves are direct representations of the course of the degradation process, then this process must have a relatively complicated course. Alternatively, if the process is assumed to follow a relatively simple course, these curves cannot give a direct representation of this course. In order to interpret the form of these curves, attention is turned to their dependence on the four experimental variables discussed below.

3.1(a) Dependence on Nominal Rate of Shear

Figure 1 shows shear load decrease plots for the degradation of ten percent *B*-100 at 40° and nominal rates of shear *D* of 66 000, 33 000, 19 000, and 9000 sec^{-1}. At the three highest nominal rates of shear concerned, the curves have the general nature already described. However, at the lowest nominal rate of shear represented the only decrease shown in the shear load is a comparatively slight one during the first few passes of the solution through the capillary, and may represent a slight thixotropic effect. At this lowest nominal rate of shear, complementary intrinsic viscosity determinations indicate that no degradation occurs.

The change in initial shear load observed in going from 9000 sec^{-1} to 19 000 sec^{-1} is comparatively very large with respect to those observed in going from 19 000 sec^{-1} to 33 000 sec^{-1}, and from 33 000 sec^{-1} to 66 000 sec^{-1}. If the initial shearing loads were determined solely by the viscoelastic resistance to flow of the solutions containing the polymer at its initial molecular weight, the relationship between initial shearing load and nominal rate of shear would not be expected to show sudden variations. The final values at which the shear load levels off after the completion of degradation under different conditions of nominal rate of shear, and of the other experimental variables to be discussed in later subsections of this paper, are determined partially by the fact that the degradation proceeds to different extents under different conditions of the experimental variables. In the present case of nominal rate of shear variation, complementary intrinsic viscosity determinations show that the higher the nominal rate of shear the greater the extent of degradation. Thus, in comparing final shear loads after degradation at two different nominal rates of shear it must be taken into consideration that the final shear load for the higher nominal rate of shear corresponds to a lower polymer molecular weight than that for the lower nominal rate of shear. In Fig. 1 the lowering of final shear load accompanying the decrease of final polymer molecular weight at a higher nominal rate of shear is not enough to overcome the increase of final shear load with increase of nominal rate of shear for a fixed polymer molecular weight. The observed result is some increase of final shear load with nominal rate of shear during degradation.

3.1(b) *Dependence on Temperature*

Figures 2 and 3 show shear load decrease plots for the degradation of 10 percent *B*-140 at 66 000 sec^{-1} and temperatures of 20°, 30°, 40°, 50°, 60°, 70°, and 80°. At temperatures of 40° and higher the curves have the general nature described earlier, but at 20° and 30° they show additional complications. At temperatures below 40° the initial shear load increases with decreasing temperature, which is an expected behavior. However, at temperatures above 50° the initial shear load increases with increasing temperature. This behavior is opposite to that which would be expected if the initial shear load were determined solely by the viscoelastic resistance to flow of solutions containing the polymer at its initial molecular weight at the experimental temperature.

This increase of initial shear load with increasing temperature above 50° is not caused by any evaporation effect. This possibility was eliminated by interrupting the run at 70° after 45 passes of the solution through the capillary, at which point the degradation had been completed and the shear load had reached its essentially constant final value. As is seen from Fig. 3, when the shearing is resumed after the 45 minute interruption, which is as long as the solution is heated before the degradation was started, no evaporation effects are detectable.

Complementary intrinsic viscosity determinations show that the higher the temperature during the degradation the smaller the extent of degradation. The fact that the final shear loads are very roughly the same for degradation at all the different temperatures indicates that the lowering of shear load by the decrease of final polymer molecular weight with decreasing temperature is roughly equal to the increase of shear load with decreasing temperature for a fixed polymer molecular weight. Whether this equivalence is significant or fortuitous is not apparent.

3.1(c) *Dependence on Concentration*

Plots of shear load decrease during the degradation of 5, 10, 15, and 20 percent solutions of *B*-100 at 40° and 66 000 sec^{-1} are shown in Fig. 4. All of these curves have the general nature described previously, and throughout the concentration range concerned the initial shear load increases with increasing concentration as would be expected.

Complementary intrinsic viscosity measurements show that the extent of degradation increases with increasing concentration. However, since the final shear loads increase with increasing concentration it is clear that the decrease of final shear load resulting from the decrease of final polymer molecular weight with increasing concentration is thoroughly overwhelmed by the normal increase of shear load with increasing concentration for a fixed polymer molecular weight.

3.1(d) *Dependence on Initial Polymer Molecular Weight*

Figure 5 shows shear load decrease plots for the degradation of 10 percent solutions of *B*-40, *B*-60, *B*-80, *B*-100, *B*-120, and *B*-140 at 40° and 66 000 sec^{-1}. These polymers show variation of initial molecular weight as indicated in Table I. All of the curves except that for the lowest initial molecular weight show the general nature described previously. The only shear load decrease shown for the lowest initial molecular weight polymer, *B*-40, is a very slight one after the first passage through the capillary. Intrinsic viscosity measurements indicate that no degradation occurs in the solution of *B*-40. In decreasing the initial molecular weight from that of *B*-140 (2 520 000) to that of *B*-80 (1 020 000), the initial shear load increases, which behavior is just opposite to that which would be expected if the initial shear load were determined solely by viscoelastic resistance to flow. However, in decreasing the initial molecular weight below that of *B*-80, the initial shear load decreases in accordance with ordinary expectations. The difference in initial loads for *B*-40 and for *B*-60 is much greater than the differences between the initial loads for any of the higher molecular weight polymers. This much greater change in going from *B*-40 to *B*-60 as compared with that in going from *B*-60 to higher molecular weight polymers is not expected on the basis of viscoelastic resistance to flow.

In general, the final shear loads decrease somewhat with decreasing initial molecular weight. Intrinsic viscosity measurements show that the final molecular weights also decrease somewhat with decreasing initial molecular weight. It might be expected that the final molecular weights would be the same for all initial molecular weights. However, the observed variation of final molecular weight with initial molecular weight can

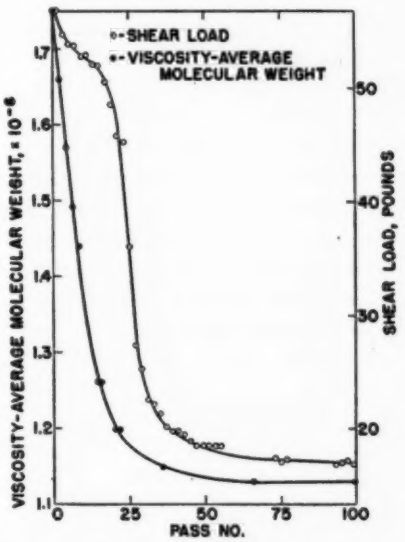


FIG. 7. Shear load and molecular weight changes during degradation of 10 percent Vistanex *B*-100 in cetane at 40° and $D = 66\,000\ sec^{-1}$.

probably be explained on the basis of the polydispersion of the molecular weights of the polymers used, and might not be observed for sharply fractionated polymers.

Figure 6 shows shear load decrease plots for the degradation of 15 percent solutions of *B*-80, *B*-100, and *B*-140 at 40° and 66 000 sec^{-1}. These curves show essentially the same type of behavior as those for the 10 percent solutions represented in Fig. 5.

3.2 Molecular Weight Decrease during Degradation

The outstanding feature shown by the comparison of the shear load decrease plot and the molecular weight decrease plot for degradation under the same conditions, as given in Fig. 7, is the correspondence between the regions during the early stages of degradation when the molecular weight is decreasing rapidly, and when the shear load is decreasing only slowly prior to its much more rapid decrease.

4. DISCUSSION

4.1 Composition of Apparent Shearing Forces

The dependence of the initial apparent shearing forces, observed at the beginning of the degradation process, on the four experimental variables discussed in the previous section is clearly not that which would be expected if these forces were associated only with overcoming viscoelastic resistance to flow. Two observed facts suggest an explanation for the observed behavior of these forces. These facts are: (1) the disproportionate decrease in initial force in going from a condition where degradation occurs to a condition where degradation does not occur, and (2) the maintenance of a relatively high value by the apparent shearing force during the stage of degradation in which the average molecular weight is decreasing rapidly, and the relatively rapid falling-off of the apparent shearing force when the molecular weight decrease becomes relatively slow. The proposed explanation of the observed behavior of the initial apparent shearing forces is that, in addition to the usual contribution associated with overcoming viscoelastic resistance to shear, they contain a contribution associated with the energy required to break the polymer molecular bonds in the degradation process.

This explanation explains the failure of the initial apparent shearing forces to behave as expected *a priori*. However, it does not make a positive explanation of the particular way in which these forces do behave. Such an explanation requires a detailed analysis of the kinetics and the energy requirements of mechanical shear degradation. An analysis of this type is being undertaken and reported in a series of papers of which reference 1 and the present paper are a part. However, this analysis is not sufficiently advanced at the present time to make any simple statement here of its implications to the behavior of the initial apparent shearing forces.

4.2 Significance to Viscometry Measurements

The fact that the apparent shearing forces observed in viscometry measurements where molecular weight and rate of shear energy input are high enough so that shear degradation occurs may not be associated exclusively with viscoelastic resistance to shear is vital to the valid utilization of such measurements. Specifically, it is necessary that the investigator not automatically assume that the observed apparent shearing forces do represent only viscoelastic resistance to shearing and that he can calculate some kind of viscosity values directly from these observed forces. In order to calculate valid viscosity values it is necessary to find some way to determine what part of the observed shearing forces do represent viscoelastic resistance to shearing, and then calculate the viscosity values from this appropriate part of the observed forces.

One procedure which has been used in the past is to assume that if one takes a reading quickly at the beginning of a viscosity measurement in which shear degradation occurs, one will thereby get a direct indication of the viscosity of the initial, undegraded material. This assumption may be entirely unjustified, since if part of the observed shearing force is associated with energy used up in the degradation process, then the force actually associated with overcoming viscoelastic resistance to shear will be much less than the observed apparent shearing force.

5. ACKNOWLEDGMENTS

The assistance of Miss H. V. Belcher and of Maria Bestul in some of the experimental work and in the preparation of figures is gratefully acknowledged. Dr. George M. Brown of the University of Maryland was in charge of the thesis from which this paper was extracted. Thanks are expressed to the Enjay Company, Inc., for supplying gratis the polyisobutene samples used.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 9 SEPTEMBER, 1954

Relaxation Time Spectrum of Dough and the Influence of Temperature, Rest, and Water Content*

J. R. CUNNINGHAM AND I. HLYNKA
Grain Research Laboratory, Winnipeg, Manitoba, Canada
(Received April 12, 1954)

The effect of temperature, rest period, and water content on stress relaxation in doughs held at constant extension has been studied with a relaxometer previously described. Temperature was varied from 13 to 35°C, rest period from 2 to 120 min, and water content from 74.4 percent to 86.0 percent dry basis. It is found that the shape of the relaxation curves on a log time plot does not change with temperature in the above range but that the curve is shifted laterally along the log time axis. This permits the construction of master curves valid in the temperature range and extends the time scale of observation. The long relaxation time end of the distribution of relaxation times is affected most by rest period, decreasing as rest period is increased, and the short time end increases as water content is decreased. Activation energies from 11 to 24 kcal per mole are calculated for relaxation from the shift of the curve on the log time axis. There is an indication that the activation energy depends on temperature and increases with rest period.

A mathematical framework, based on the general Maxwell model, is discussed and the distribution function on the log time axis is interpreted in terms of viscosity and modulus of the relaxing mechanisms. Special reference is made to the "box distribution" and a viscosity is calculated.

SCOTT-BLAIR and others[1–5] have conducted a number of experiments with the aim of measuring elasticity and viscosity of doughs made from wheat flour. Work in this Laboratory[6,7] has indicated that elasticity and viscosity of dough cannot be defined in the classical manner since relaxation is not Maxwellian. However, the relaxation behavior may be described by considering an array of Maxwell elements resulting in a distribution or spectrum of relaxation times. A study of relaxation becomes a study of the distribution of relaxation times, and the state of a dough must be represented by curves rather than single values.

In the present investigation a method, known as stress relaxation, that has been developed in the field of polymer science has been applied to the study of dough properties.[6] Tension in the relaxing dough is plotted against the logarithm of time. This yields a relaxation curve which is used as the basis for study. An approximation to the distribution of relaxation times in the substance is obtained graphically from it and the distribution is examined for different dough treatments.

Temperature has a marked effect on the relaxation process in dough, but it is found that the shape of the relaxation curve is not greatly altered, rather that the curve is shifted laterally along the logarithmic time axis. This is a well-known result in polymer studies.[8–10] The curves taken at different temperatures may therefore be combined to construct a master or composite curve[11] and thus to extend the time range of observation beyond any one curve. The master curve is valid at any temperature within the range examined.

Master curves for dough and distribution functions obtained from them are presented to show the effect of resting and of water content.

Apparent activation energies are calculated from the magnitudes of the temperature shifts.[11,12] The indication is that an Arrhenius activation energy can be used to characterize the relaxation process and that it changes only slightly with rest and water content within the range examined.

In the discussion, theory is included which deals with a mathematical basis for the treatment of the relaxing process.

METHOD AND MATERIALS

Measurements of relaxation of tension were made using an instrument called a relaxometer which has been described in a previous paper.[6] This instrument stretches a ball of dough to a preset distance, holds the sample at a constant extension, and records the way in which the tension relaxes with time.

Straight-grade flour of 13.1 percent protein milled on a laboratory mill from spring wheat grading No. 3 Northern was used. Calculated proportions of flour and water were used to give a dough weighing 150 g which

* Paper No. 134 of the Grain Research Laboratory, and No. 314 of the Associate Committee on Grain Research (Canada).

[1] R. K. Schofield and G. W. Scott-Blair, Proc. Roy. Soc. (London) **A138**, 707 (1932).

[2] R. K. Schofield and G. W. Scott-Blair, Proc. Roy. Soc. (London) **A139**, 557 (1933).

[3] R. K. Schofield and G. W. Scott-Blair, Proc. Roy. Soc. (London) **A141**, 72 (1933).

[4] R. K. Schofield and G. W. Scott-Blair, Proc. Roy. Soc. (London) **A160**, 87 (1937).

[5] M. P. Volarovich and R. A. Branopolskaya, Kolloid Zhur. **10**, 406 (1948).

[6] Cunningham, Hlynka, and Anderson, Can. J. Technol. **31**, 98 (1953).

[7] I. Hlynka and J. A. Anderson, Can. J. Technol. **30**, 198 (1952).

[8] T. Alfrey, Jr., *Mechanical Behavior of High Polymers* (Interscience Publishers, Inc., New York, 1948), p. 132.

[9] C. A. Dahlquist and M. R. Hatfield, J. Colloid Sci. **7**, 253 (1952).

[10] F. Schwarzl and A. J. Staverman, J. Appl. Phys. **23**, 838 (1952).

[11] Andrews, Hofman-Bang, and Tobolsky, J. Polymer Sci., **3**, 669 (1948).

[12] A. V. Tobolsky and J. R. McLoughlin, J. Polymer Sci. **8**, 543 (1952).

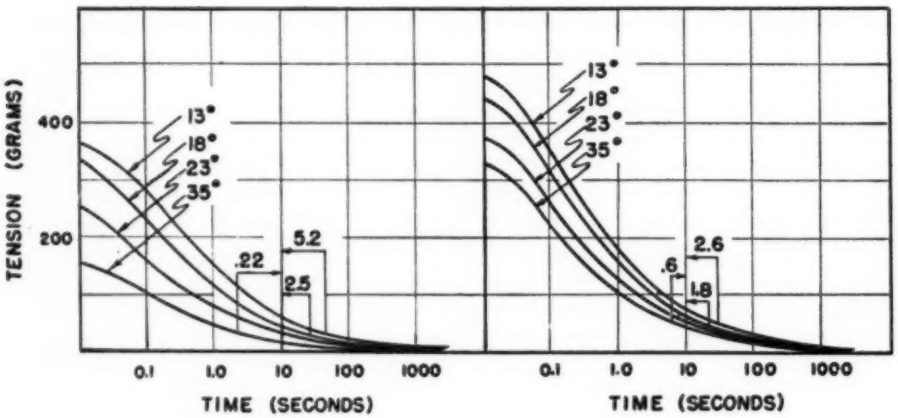


FIG. 1. Effect of temperature on stress relaxation in dough. (a) Water content 86.1 percent (dry basis) and 120 min rest. (b) Water content 80.3 percent (dry basis) and 120 min rest.

was mixed for 3 minutes in a laboratory mixer.[13] Immediately after mixing the dough was moulded mechanically into a flattened ball and impaled upon a split pin dough holder[6] to form a sort of ring or doughnut. The sample was then conditioned for a time which is termed rest period[6,7] at a preset temperature and a relative humidity of 95 percent.

The relaxation part of the curve is replotted on semi-log paper as tension *versus* the logarithm of the time and the negative of the slope of this curve is found graphically and plotted against $\log\tau$ where τ is the relaxation time and replaces t, the time since extension ceased.

Tension rather than stress is used as the ordinates of the curves shown since the cross-sectional area is difficult to measure accurately; it does not change during relaxation and it will be the same in each case because of the constant moulding and stretching procedure.

EXPERIMENTAL RESULTS

Four temperatures were used, and where possible 4 rest periods and 4 water contents. At least two curves were drawn for each setting and where reproducibility was not satisfactory more were drawn. In all some 150 doughs were examined.

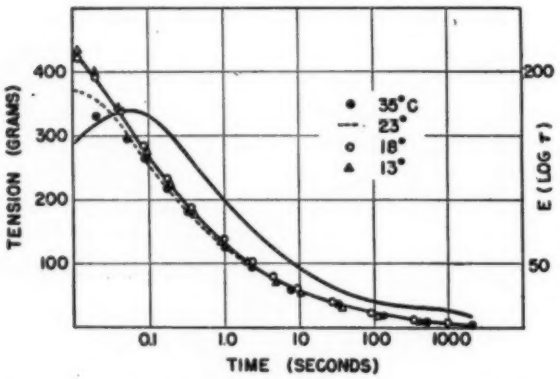


FIG. 2. Master relaxation curve and distribution curve for stress relaxation in dough. Water content 80.3 percent (dry basis), rest period 120 min.

[13] Dempster, Hlynka, and Anderson, Cereal Chem. **30**, 492 (1953).

Of necessity, all these data cannot be presented but representative portions are given in a summarized form by the use of master curves.

I. The Effect of Temperature and the Master Curves

The effect of temperature on the relaxation curves for dough is shown in Figs. 1(a) and 1(b), where curves for water contents of 86.1 percent and 80.3 percent dry basis (60 percent and 55 percent wet basis or bakers absorption) are drawn. The rest period used was 120 minutes.

A large number of investigators[11,12,14–17] have shown that for many polymers the relaxation curves taken at different temperatures superpose if the time values at one of the temperatures are all multiplied by a certain factor. This is equivalent to a lateral shift on the log time axis. It is found that the curves shown in Fig. 1 also bear this relation to each other. The amounts of the shift are indicated on the diagram. The quantitiee shown are the ratios of two τ values or the "tims factors"[11,12] at different temperatures.

The fact that the shape of the relaxation curve is not altered by a temperature change permits the construction of a master curve at one temperature that is valid for any temperature within the range involved. This has the effect of extending the time scale of observation. The master curve for the curves in Fig. 1(b) is shown in Fig. 2. The dashed curve was taken at 23°C. It is in mid range and was selected as a reference to which the other curves were fitted. The filled circles are from 35°C, the open circles from 18°C and the triangles from 13°C. The solid line is the composite or master curve. The deviations of the individual curves from the master curve are slight except for the region of relaxation times less than one-tenth second. The finite time of extension

[14] R. D. Andrews and A. V. Tobolsky, J. Polymer Sci. **7**, 221 (1951).
[15] Bischoff, Catsiff, and Tobolsky, J. Am. Chem. Soc. **74**, 3378 (1952).
[16] G. M. Brown and A. V. Tobolsky, J. Polymer Sci. **6**, 165 (1951).
[17] J. D. Ferry, J. Am. Chem. Soc. **72**, 3746 (1950).

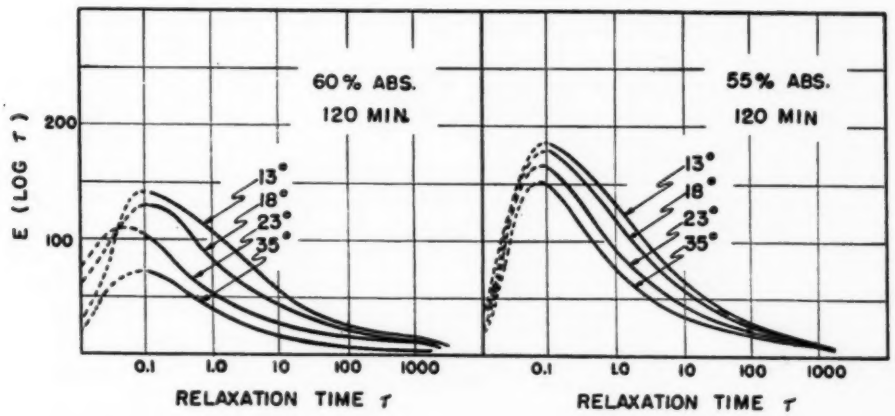


FIG. 3. Relaxation time distributions for curves shown in Fig. 1.

has a strong effect on this region in that it causes the curve to level off where it otherwise would not.

In comparing the curve at one temperature to that at another a small adjustment in the stress values should also be made[12] since, accordlng to the kinetic theory of elasticity, the stress should be proportional to the absolute temperature. The temperature range studied is rather limited and this correction is within the experimental error and so has been neglected.

The distributions of relaxation times for the curves in Figs. 1(a) and 1(b) are shown in Figs. 3(a) and 3(b). They are obtained graphically by plotting the negatives of the slopes of the curves in Figs. 1(a) and 1(b) against log$t$. The distributions of relaxation times are also shifted to shorter relaxation times as the temperature is increased. There appears to be a certain variation in their shapes but this is not systematic and is due in part to the difficulty in graphically obtaining accurate slopes of the relaxation curves. The low relaxation time part of the distribution curve is dotted from the point where it begins to fall off to indicate that in this region the true distribution is not obtained. The distribution curve for the master curve is shown in Fig. 2.

II. Effect of Rest Period

Rest period is the time interval between moulding the dough (after mixing) and stretching it on the relaxometer.

The effect of rest period is shown by the relaxation curves in Fig. 4(a), where curves are shown for doughs having rest periods of 2, 5, 45, and 120 minutes. Only 3 curves appear as the curve for 120 minutes falls almost exactly on that for 45 minutes rest. The water content was 80.3 percent dry basis (or 60 percent absorption). The triangles and circles, etc., indicate points from the individual curves. The corresponding distribution curves are shown in Fig. 4(b).

The unrested dough (2 min rest) shows a broad distribution of relaxation times which could be approximately represented by a "box distribution"[11,14] as is shown in Fig. 4(b). As the dough rests, the part of the curve associated with relaxation times greater than of the order of 10 seconds shows quite a marked decrease and the "box" approximation becomes poorer. This effect is quite prominent at all water contents tried, although it is most pronounced when water content is low (absorptions of 50 and 45 percent).

That dough properties change during resting is well known to bakers and it is called "slackening" by them. A more scientific term, structural relaxation, has been suggested.[13] From the way in which the distribution curves change it might be inferred that this effect is associated with mechanisms which have relaxation times greater than 10 seconds. The change is time dependent and is very nearly complete in 45 minutes, i.e., a dough rested longer than this is said to have "stabilized." This is illustrated by Fig. 4 where the 45- and 120-minute curves superpose.

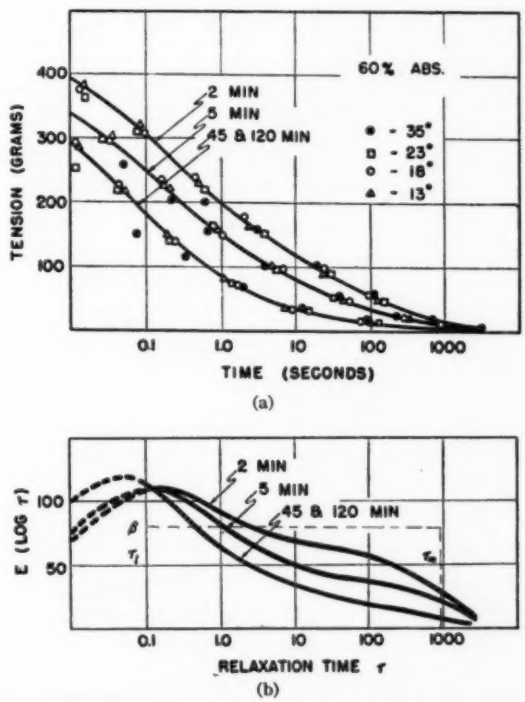


FIG. 4. Effect of rest period. (a) Master relaxation curves. (b) Distribution curves.

The process of stabilization or "structural relaxation" is temperature dependent and progresses more rapidly at higher temperatures. Therefore it is necessary, in comparing doughs that have been rested for short periods, that the states of "slackening" be comparable. It was found that the curve for a dough rested 5 minutes at say, 23°C, did not superpose on the curve for a dough rested 5 minutes at 13°C. Rather, at 13°C a rest period of approximately 13 minutes should be used. Similarly, at 35°C the dough should be rested for only 2 to 3 minutes. When these adjustments were made curves were obtained that produced a fairly good master curve. The adjustment was made in some cases by testing new doughs and in some cases by interpolation. The amount of adjustment required becomes less as rest period is increased and at 120 minutes none at all is needed.

III. Effect of Water Content

The effect on the master relaxation curves of a change in water content is shown in Fig. 5(a) where curves are drawn for a rest period of 120 minutes and water contents of 74.4 percent, 80.3 percent, and 86.0 percent dry basis. This is 50 percent, 55 percent, and 60 percent bakers absorption. The corresponding distribution curves are shown in Fig. 5(b). The greatest change now appears to be in the short relaxation time part of the distribution. The ordinates over the whole range of the curve are increased when water content is low but those associated with a τ value of less than 10 seconds are increased most. The changes are similar at all rest periods and are even more marked in unrested doughs.

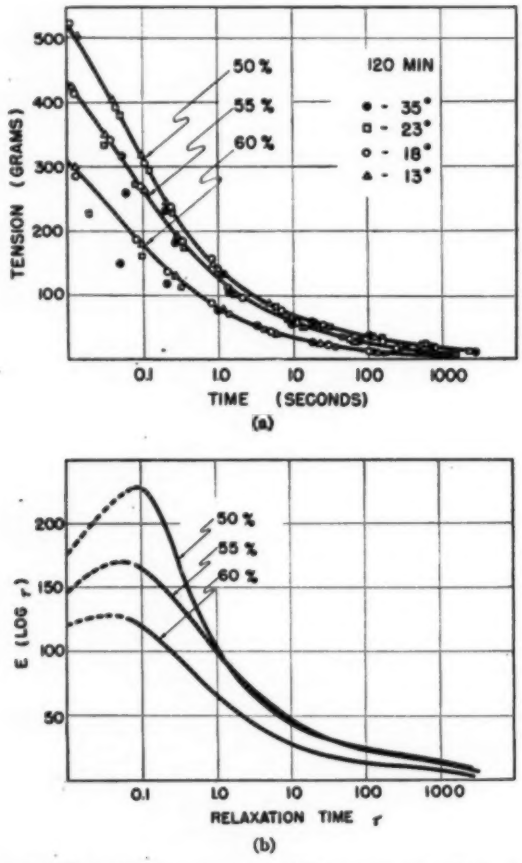


FIG. 5. Effect of water content. (a) Water contents 74.4 percent, 80.3 percent, and 86.0 percent dry basis (50 percent, 55 percent, and 60 percent wet basis). (b) Distribution curves.

This and unpublished data from this Laboratory suggest that the effect of water on relaxing dough is connected most with those mechanisms which have short relaxation times.

IV. Activation Energy

The simple behavior of the relaxation curves with temperature permits the calculation of an Arrhenius energy of activation for stress relaxation.[11,16,17] It is assumed that all the relaxing mechanisms have the same temperature dependence so that each may be represented by an Arrhenius equation.[9]

$$\tau_i = A_i e^{E/RT}, \tag{1}$$

where τ_i is the relaxation time of the ith mechanism at temperature T, and A is a constant evaluated at a reference temperature. The behavior of τ_i with temperature is then given by the shift of the relaxation curve along the log time axis. The activation energies are obtained from the slopes of the curves by plotting the logarithm of the shift against the reciprocal of the absolute temperature.

Figure 6 shows the curves obtained for 3 absorptions and 4 rest periods. Because of a limited supply of flour of this exact type, and the fact that protein denaturation is likely to take place at temperatures much higher than those examined, only 4 points on each curve were available.

The scatter in the values obtained is considerable but serves to indicate the order of the apparent activation energy. The values obtained are presented with the curves. There is a suggestion of an increase with rest period, especially at high water content; however, more extensive and more accurate data would be required to assess this change. The values obtained are from 11 to 24 kcal and may be compared with 14–23 for polyisobutylene[9,11,12,14] and 15–28 kcal for GR-S[9] and values for a number of substances given by Morey.[18]

The activation energies calculated for relaxation are slightly higher than those calculated from viscosity data. Fox and Flory show E less than 10 for polyisobutylene for this temperature range.[19]

The energies are temperature dependent, being somewhat higher for lower temperatures. This is also a common finding and will be dealt with further in the discussion.

[18] D. R. Morey, J. Phys. & Colloid Chem. **53**, 569 (1949).
[19] G. J. Dienes, J. Appl. Phys. **24**, 779 (1953).

THEORETICAL CONSIDERATIONS AND DISCUSSION

The first two papers[6,7] of this series dealing with the decay of stress in dough at constant extension present data to show that dough is not a simple Maxwellian substance but should be considered in terms of a more general model. These data together with those in the present paper are now sufficient in scope to warrant a discussion of the general background of stress relaxation in dough. Accordingly, a mathematical description of the relaxation curves based on a generalized Maxwell model is considered at some length. Allowance is made for the finite time of extension and an interpretation of the relaxation data in terms of modulus and viscosity is sketched. Reference is made to a "box distribution" which obtains from a system in which the viscous elements are all the same. The activation energies as found from relaxation data are also discussed briefly.

I. The Relaxation Curve and the Distribution Function

Dough, along with a number of other substances, exhibits a complex behavior in stress relaxation which may be represented by a spectrum of relaxation times. Such a representation serves to preserve the ideas of viscosity and elasticity and linear behavior and yet extend them to deal with more complex results.[20] The relaxation times are considered to arise from a continuous array of Maxwell elements, the moduli of which may be characterized by a distribution function $E(\tau)d\tau$ which is defined as the total modulus from those elements having relaxation times between τ and $\tau+d\tau$.

The relaxation of stress is then described by an integral,

$$S(t)=\gamma\int_0^\infty E(\tau)e^{-t/\tau}d\tau, \tag{2}$$

where $S(t)$ is the stress at time t and γ is the strain.

An easy substitution, $1/\tau=s$ shows that $E(\tau)$ bears a very simple relationship to the inverse Laplace transform of $S(t)$.[21] Since $S(t)$ is not known analytically one must use approximate methods to obtain $E(\tau)$ from $S(t)$. The integral (2) may be expanded as a limit of a series,[22] the first term of which is

$$E(\tau)=-\frac{d}{dt}[S(t)]_{t=\tau}. \tag{3}$$

The experimentally more useful form of this is in terms of the logarithm of time

$$E(\ln\tau)=-\frac{d}{d(\ln\tau)}[S(t)]_{t=\tau}. \tag{4}$$

[20] B. Gross, J. Appl. Phys. **18**, 212 (1947).
[21] F. Schwarzl and A. J. Staverman, Physica **18**, 791 (1952).
[22] G. Doetsch, *Theorie und Anwendung der Laplace Transformation* (Dover Publications, New York, 1943).

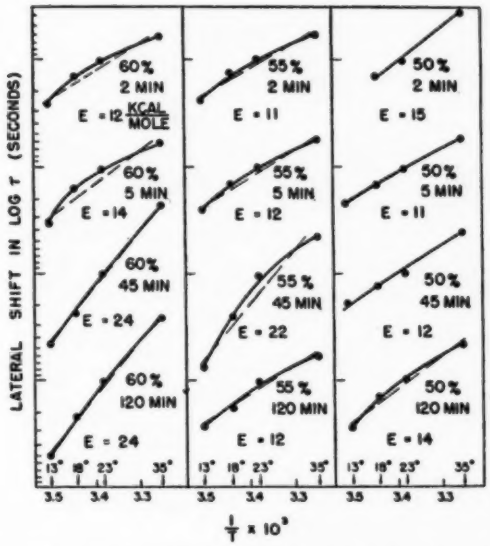


FIG. 6. Activation energies. Plot of lateral shifts (time factors) against reciprocal absolute temperature.

This first-term approximation is frequently referred to as the Alfrey approximation[23] and is widely used in the literature. It may be obtained graphically by measuring the slopes of the relaxation curve and plotting their magnitude against the logarithm of time. Second-order approximations are occasionally reported[21,24] but only one term is considered in the present work. Plotting the relaxation curves this way does not add new information, but it does put the existing data into a more usable form.

II. The Effect of Extension Rate

The extension process requires about 1 second and since this is comparable to the first part of the relaxation curve many of the short relaxation time mechanisms likely relax during the extension and do not appear in the observed spectrum. That this is so may be shown as follows.

The sample is strained at a constant rate to an amount γ in a time t_0 so that the rate of strain is γ/t_0. The stress built up by this process in a generalized Maxwell model is given by[25]

$$S(t_0)=\frac{\gamma}{t_0}\int_0^\infty \tau E(\tau)(1-e^{-t/\tau})d\tau, \tag{5}$$

which may be written as

$$S(t_0)=\gamma\int_0^\infty E(\tau)\left[\frac{1-e^{-t/\tau}}{t_0/\tau}\right]d\tau. \tag{6}$$

If the extension is stopped at t_0 and relaxation is measured from that time on, the relaxation may be

[23] T. Alfrey, Jr. and P. M. Doty, J. Appl. Phys. **16**, 700 (1945).
[24] J. D. Ferry and M. L. Williams, J. Colloid Sci. **7**, 347 (1952).
[25] T. Alfrey, Jr., *Mechanical Behavior of High Polymers* (Interscience Publishers, Inc., New York, 1948), p. 186.

described by

$$S(t,t_0)=\gamma\int_0^\infty E(\tau)\left[\frac{1-e^{-t_0/\tau}}{t_{0/\tau}}\right]e^{-t/\tau}d\tau. \qquad (7)$$

The effect of the time of extension is thus a modification of $E(\tau)$ by the term in square brackets. This term is very nearly zero when τ is less than t_0 and is very nearly unity when τ is greater than t_0. Ter Haar[26] has replaced this function by what he calls a "cutting-off function"

$$S(t_0,t)=\gamma\int_0^\infty E(\tau)C(t_0/\tau)e^{-t/\tau}d\tau$$
$$=\gamma\int_{t_0}^\infty E(\tau)e^{-t/\tau}d\tau. \qquad (8)$$

This indicates that the part of the distribution that has relaxation time τ less than t_0, the time of extension, is not observed. This region of the spectrum could be studied best by dynamic methods.

This fact must be borne in mind when interpreting the distribution curve data and is the reason for dotting the first part of the curves.

III. Interpretation of the Relaxation Data in Terms of Modulus and Viscosity

It is possible to interpret the distribution curves in terms of viscosities and moduli of the relaxing mechanisms and to get a value for the total viscosity.

The distribution function $E(\tau)d\tau$ was defined earlier as the modulus for those elements that have relaxation times between τ and $\tau+d\tau$. This may be seen to be an element of area under the distribution curve. The total Young's modulus would then be the area under the complete curve.

$$E_{\text{total}}=\int_0^\infty E(\tau)d\tau=\int_{-\infty}^{+\infty}E(\ln\tau)d(\ln\tau). \qquad (9)$$

On the log plot finite limits must be taken to prevent the integral from diverging. Only an apparent modulus can be obtained from the present data since the short τ end of the spectrum is not determined.

The viscosity of an interval $d\tau$ of the distribution is $\eta(\tau)d\tau=\tau E(\tau)d\tau$ which is the moment of the area under the curve. The total "tensile" viscosity is then

$$\eta_{\text{total}}=\int_0^\infty \tau E(\tau)d\tau=\int_{-\infty}^{+\infty}\tau E(\ln\tau)d(\ln\tau). \qquad (10)$$

Again, only an apparent viscosity can be obtained although the total viscosity is more dependent on the long τ end of the spectrum than is modulus.

It can be shown that $E(\ln\tau)$, the ordinate at τ of the distribution curve on the semi-log diagram, represents the viscosity of the mechanism with that τ value.

Since $E(\tau)d\tau=E(\ln\tau)d(\ln\tau)$, then $E(\ln\tau)$, the distribution on the logτ plot is equal to $\tau E(\tau)$, and if we take $d\tau$ equal to 1 as does Kuhn[27] then this is numerically equal to the viscosity of the mechanisms with relaxation time τ to $\tau+1$. The ordinates of the distribution curves thus represent the distribution of the viscosities of the viscous elements. The distribution of the elastic elements may be obtained from this by dividing the ordinates by τ.

The data for the effect of rest period may be considered in this light. Figure 4 shows that the viscosities associated with Maxwell elements that have relaxation times greater than 1 second are decreased when the dough rests while the viscosities of those elements that have relaxation times shorter than this are affected to a much less degree.

IV. A Box Distribution

A number of authors report relaxation curves which are linear over a considerable region on a log time plot.[6,7,11,14,28] This is characterized conveniently by assuming a so-called "box distribution," Fig. 4(b), which is constant with respect to logarithmic time between certain limits. This is written as

$$E(\ln\tau)=\beta, \quad \tau_1<\tau<\tau_m. \qquad (11)$$

The modulus for the ith element of this distribution is

$$E_i=E(\tau_i)\Delta\tau_i=\frac{\beta}{\tau_i}\Delta\tau_i=\frac{\beta}{\tau_i}. \qquad (12)$$

The viscosity of the ith element is

$$\eta_i=\tau_iE(\tau_i)\Delta\tau_i=\beta\Delta\tau_i=\beta, \qquad (13)$$

β, the height of the box is the viscosity which is associated with each relaxing mechanism and is the same throughout. This indicates that there is a distribution of moduli but only a single viscosity. The total viscosity is $\beta\tau_m$[14], where τ_m is the upper limit to the box. Here it is assumed that $\tau_m\gg\tau_1$.

β is the slope of the relaxation curve which in this case is a constant. Reference to Fig. 4 shows that the distribution curve for dough rested for very short period can be approximated fairly well by a box distribution. The approximation becomes poorer as the dough is rested. It could, however, for purposes of comparison, be represented by two box distributions: one bounded on the left by $\tau=0.1$, limited by the extension process, and on the right by $\tau=10$; and a second from $\tau=10$ to the limit of the curve. The areas under these could be measured and studied as rest period is changed as a means of quantitatively assessing the changes. This has

[26] D. Ter Haar, Physica 16, 738 (1950).

[27] W. Kuhn, Makromol. Chem. 6, 224 (1951).

[28] A. V. Tobolsky, J. Am. Chem. Soc. 74, 3786 (1952).

TABLE I. Rigidity moduli, from area.

	from $\tau=0.1$ to $\tau=10$ sec	from $\tau=10$ to limit	Apparent modulus
60 percent abs			
2 min rest	0.65×10^4 (dynes/cm²)	0.42×10^4 (dynes/cm²)	1.07×10^4 (dynes/cm²)
5 min rest	0.58×10^4 (dynes/cm²)	0.29×10^4 (dynes/cm²)	0.87×10^4 (dynes/cm²)
45 min rest	0.49×10^4 (dynes/cm²)	0.12×10^4 (dynes/cm²)	0.61×10^4 (dynes/cm²)
120 min rest	0.49×10^4 (dynes/cm²)	0.11×10^4 (dynes/cm²)	0.60×10^4 (dynes/cm²)
120 min rest			
50 percent abs	0.83×10^4 (dynes/cm²)	0.19×10^4 (dynes/cm²)	1.02×10^4 (dynes/cm²)
55 percent abs	0.73×10^4 (dynes/cm²)	0.18×10^4 (dynes/cm²)	0.91×10^4 (dynes/cm²)
60 percent abs	0.49×10^4 (dynes/cm²)	0.11×10^4 (dynes/cm²)	0.60×10^4 (dynes/cm²)

been done for the curves presented in Figs. 4(a) and 5(b), and the results are given in Table I. These results could be interpreted, bearing in mind the argument that the area under the distribution curve totals the apparent modulus (Eq. 9). The total apparent rigidity moduli are also given in Table I and the values are of the same order but slightly smaller than those calculated by Halton and Scott-Blair[29] and within the range of those reported by Halton.[30]

To get a viscosity in this manner one would have to approximate as a box distribution and obtain a limit for the spectrum. This has been done for the curve shown in Fig. 4(a) for a dough of 60 percent absorption and 2-minute rest period. The box distribution is shown in Fig. 4(b). The calculation was made by drawing the best straight line for the relaxation curve, measuring its slope and extrapolating it to the time axis. β is given by 2.303 times the slope. The ordinates are put in terms of stress by converting to dynes and dividing by the cross-sectional area, which is rather difficult to measure accurately but is very nearly 22 cm². τ_m is obtained[14] by multiplying the intercept by 1.78. The value for β is about 8.4×10^3 dynes and for τ_m is about 900 seconds. This gives 2.5×10^6 poises for the bulk viscosity, which may be compared with values of from 5×10^6 to 24×10^6 poises given by Schofield and Scott-Blair[2,3] for relaxation data on dough and from 1.6×10^6 to 11×10^6 poises reported by Halton.[30] It would be more difficult to calculate a viscosity in this way for rested doughs since the box distribution is not readily applicable to them.

The data presented in paper 1 of this series,[7] in which a parameter K was used could now be interpreted in this light, for $\beta=23.03K$.

[29] P. Halton and G. W. Scott-Blair, J. Phys. Chem. **40**, 561 (1936).

[30] P. Halton, Cereal Chem. **26**, 24 (1949).

V. Activation Energies

The activation energies evaluated in this short temperature range are presented mainly to indicate the order of magnitude and the trend of them. The dependence on temperature is in some cases slight but in others considerable. It is, however, systematic and indicates that the activation energies are lower at higher temperatures. The temperature dependence appears to be greatest at high water content and for unrested doughs. Dienes, in a recent paper,[19] describes a temperature dependence for activation energies found from pure viscosity measurements which is in the same direction but over a much greater range. This is attributed to a "structure activation energy" which is connected with the state of order in the material and the configurational changes that would be associated with flow. It is likely that this could also be considered in terms of entropy, as does Eyring,[31] and that this contribution to the activation energy for dough would be appreciable. No attempt is made to evaluate the entropy however.

The foregoing mathematical analysis helps in the understanding of the relaxation phenomenon in dough and will also form the basis of future work. It is based on the generalized Maxwell model and has therefore the limitations inherent in that assumption. The mechanisms in the theoretical model must, of course, have their counterpart in the structure of the actual dough system. So far the Maxwell model has been the one most successfully applied and should be exploited to its limit. But an alternate approach, which is not based on a mechanical model, has been introduced by Eyring and also may well form a basis for further work.

Physical investigations on dough supplemented with complementary chemical studies promise significant achievements in the future.

[31] Glasstone, Laidler, and Eyring, *The Theory of Rate Processes* (McGraw-Hill Book Company, Inc., New York, 1941), pp. 477–516.

Constant Stress Elongation of Soft Polymers: Plasticizer Studies

MARSHALL R. HATFIELD AND GEORGE B. RATHMANN
Central Research Department, Minnesota Mining and Manufacturing Company, St. Paul, Minnesota
(Received April 12, 1954)

Constant stress measurements have been made on lightly cross linked GR-S and polyisobutylene of 7 million molecular weight. The modulus values as a function of temperature, time, and concentration of mineral oil plasticizer have been treated according to the "reduced variables" principles.

Since both plasticized and unplasticized samples showed the expected equivalence of temperature and time, the modulus values could be shifted to produce master curves for each composition. In the case of GR-S the concentration of plasticizer could be similarly reduced to yield a single curve for all the data.

The polyisobutylene results, however, showed quite clearly that the plasticizer had a unique effect not paralleled by temperature or time. This effect has been attributed to the pseudoequilibrium modulus which apparently results from chain entanglements. Since the pseudoequilibrium modulus should be proportional to the number of chain entanglements per gram of polymer, it should vary as the square of the volume fraction of polymer. It is shown that a correction of this type improves the shift of the curves for polyisobutylene.

Changes in activation energy with the addition of plasticizer are apparently within experimental error (±2 kcal) in all cases.

INTRODUCTION

THE equivalent influence of time and temperature on the elastic properties of certain rubbery polymers has been demonstrated by a number of investigators.[1–4] Recently, Ferry[2] showed that dynamic rigidity-log frequency data for polymer solutions might be treated in a similar "reduced variables" manner.

Although Ferry has pointed out the limitations of such a treatment particularly for solutions of high polymer concentration, the method is still intriguing as an approach to a plasticizer treatment. This investigation represents an attempt to describe the conditions under which the reduced variable treatment might be expected to apply for elastomer-plasticizer systems.

THEORETICAL

Elastomer deformation behavior can be represented by plotting the elastic modulus (stress-strain ratio) against logarithm of time giving, ideally, a curve like the solid portion of Fig. 1.

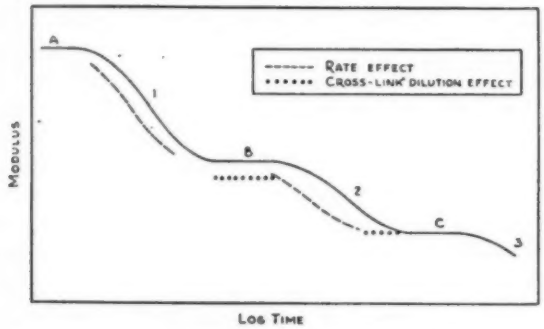


FIG. 1. Hypothetical modulus curve.

Unchanging, or "equilibrium," portions of the curve are a result of "cross links" which give rise to a network structure. The network may be permanent, in which case the curve will remain flat or it may be temporary which means, of course, the modulus will drop further after a suitable length of time.

In Fig. 1, for example, the processes which cause modulus changes (so-called dispersion regions) might be as follows:

Process 1—Between A and B: Slippage of chain segments which were held together at A by secondary valence forces—characterized by relaxation time τ_1.

Process 2—Between B and C: Slippage of longer units, perhaps held together by mere physical entanglements—characterized by relaxation time τ_2.

Process 3—Beyond C: Viscous flow or molecular slippage brought about by chemical rupture of primary valence bonds—characterized by relaxation time τ_3.

Addition of plasticizer to the elastomer will, in general, have three predominate effects. These are (a) chain dilution effect, (b) rate effect, and (c) "cross link" dilution effect. The chain dilution effect is the decrease in concentration of polymer per unit volume upon plasticization. The whole curve is corrected for this effect through multiplication by a constant factor. For this paper, modulus values of plasticized material were calculated on the basis of the swollen cross section and swollen initial length. A multiplying factor of $v^{-\frac{1}{3}}$ (v=volume fraction polymer) will correct the curves for chain dilution which will mean that all values are properly based on the unswollen cross section.

The rate effect is a result of changing the internal viscosity which, of course, changes the relaxation time of any given process. The influence of plasticizer on rate will depend upon the process; that is, one process is affected in a manner different from another. For a

[1] R. D. Andrews and A. V. Tobolsky, J. Chem. Phys. **13**, 3 (1945).
[2] J. D. Ferry, J. Am. Chem. Soc. **72**, 3746 (1950).
[3] H. Leaderman, *Elastic and Creep Properties of Filamentous Materials* (The Textile Foundation, Washington, D. C., 1943).
[4] C. A. Dahlquist and M. R. Hatfield, J. Colloid Sci. **7**, 253 (1952).

given process which is characterized by a distribution of relaxation times, however, it seems likely that the influence of plasticizer will be constant. The plasticizer effect would then be manifested by a shift of the dispersion portion of the modulus curve along the log time axis (see broken curves in Fig. 1).[2] The amount of curve shift for process 2 in Fig. 1 can be calculated from measurements of steady-flow viscosities of pure and plasticized polymer. If there are no physical chain entanglements (B) or primary cross links (C), processes 1 and 3 will merge, in which case the shift for most of the curve might be closely approximated by viscosity determinations. In the absence of primary cross links, this is probably the situation for very high plasticizer (or solvent) concentrations.

The effect of plasticizer on cross links will depend upon both plasticizer and type of cross link. If the cross links are a result of primary valence forces the plasticizer probably will have little effect. On the other hand, a good plasticizer will reduce the concentration of "entanglements" (see dotted lines in Fig. 1) and may reduce the number of secondary valence cross links. Until region A is more completely defined, however, it will be assumed that plasticizer has no effect on the "equilibrium" value of the "glass" modulus.

TABLE I. Slopes of GR-S:mineral oil creep curves (25°C).

Composition, % oil	Slope ($d \log t / d \log G$), min per megadyne
0	−5.27
13	−5.14
23	−5.33

EXPERIMENTAL

The constant stress apparatus has been described in detail in earlier papers.[4,5] The polymer sample which was cut from a thin film is loaded with a hyperboloid shaped weight which sinks into a liquid as the film stretches. By suitable design, the load is reduced in proportion to the change in cross sectional area of the stretching film.*

The polyisobutylene, which was furnished by John Rehner, Jr., of Esso Laboratories, has a molecular weight of approximately 7×10^6. It is soluble in heptane so test films could be deposited from solution. When plasticizer was used, it was added to the solution before the film was cast. Further details of the film-casting technique are described in the earlier papers.

The GR-S was prepared in this laboratory by S. Smith. Polymerization of a 70:30 butadiene:styrene charge was carried out using a low modifier concentration such that the polymer exhibited a gel value of 70 percent and swell ratio of 30 (benzene). The polymer was milled briefly and then dispersed in toluene in an internal mixer. Plasticizer was added to the mixture in the mixer. Sample films were deposited from the toluene dispersion to a thickness of about 0.020 in.

The modulus values were calculated as the stress-strain ratio and reduced for temperature in the conventional manner.

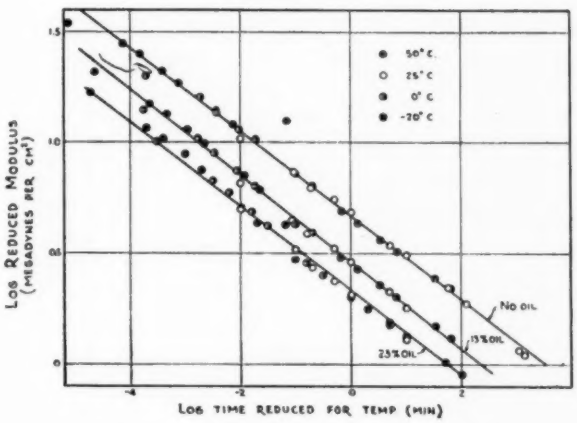


FIG. 2. Creep curves: GR-S:mineral oil system.

RESULTS AND DISCUSSIONS

Only two systems have been studied so the results do not provide a complete picture. They do, however, provide information concerning two portions of the idealized curve in Fig. 1. Creep curves for the GR-S:mineral oil system at 25°C are given in Fig. 2. Several temperatures (−20°C, 0°C, 25°C, 50°C) were involved and the data reduced to 25°C by the curve shifting method.† Even though the polymer was lightly cross linked during polymerization (approximately 70 percent gel, 30 swell ratio in benzene), the curve shows no approach to an equilibrium value. This is not surprising since this value would probably be less than 1×10^6 dynes per square centimeter. Note that the curve is well represented by a straight line on the log-log plot over seven decades of logarithmic time. Apparently there are no physical chain entanglements giving rise to a pseudo-equilibrium modulus like region B in Fig. 1.

The slopes, as calculated by the least-squares method, of the GR-S:mineral oil curves are given in Table I. Since the greatest difference is only about 4 percent and, furthermore, since there is no regular variation, it will be assumed that the curves superpose nicely by a shift along the log time axis. This means that the plasticizer had merely a rate effect on the polymer behavior which implies that the oil was not effective in changing equilibrium modulus values, or in other words, that the change in the rate of the deformation process was due entirely to a change in free energy of activation. This energy change which would be necessary to cause a

[5] Dahlquist, Hendricks, and Taylor, Ind. Eng. Chem. 43, 1404 (1951).

* The assumption is made that the constant stress treatment remains valid in the region of a pseudo-equilibrium modulus.

† For a description of the "reduced variables" method, see references 1, 2, 4, 6.

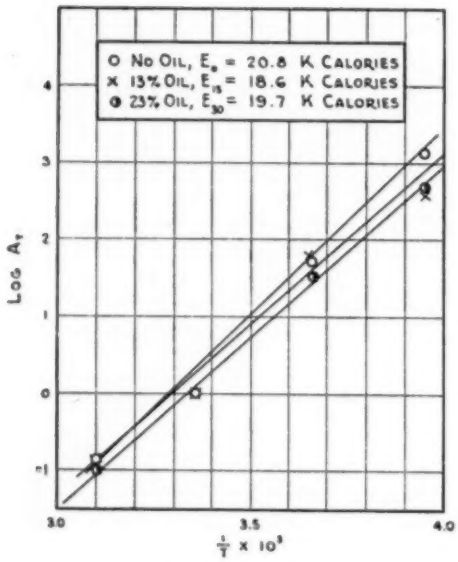


FIG. 3. Temperature shift factors: GR-S:mineral oil system.

given rate increase is determined by

$$\Delta F^{\ddagger} = F_1^{\ddagger} - F_0^{\ddagger} = 2.303RT \log(t_1/t_0), \qquad (1)$$

where t_0 and t_1 are the times required to attain a given modulus value for pure polymer and plasticized polymer, respectively. Using GR-S data for the 13-percent oil composition, $\Delta F^{\ddagger}$ was calculated to be 1.4 kcal. This change may be a result of a change in heat of activation (so-called "activation energy") or a change in entropy of activation or both. Since it is difficult to evaluate entropy of activation, it will be assumed that the measured rate change was entirely a result of "activation energy" change. The apparent "activation energy" for a given system might be calculated from the temperature dependence of the creep curves. In Fig. 3, the temperature shift factors for each GR-S:mineral oil composition are plotted against reciprocal temperature. The data are best represented by straight lines, the slopes of which are equal to apparent activation energy

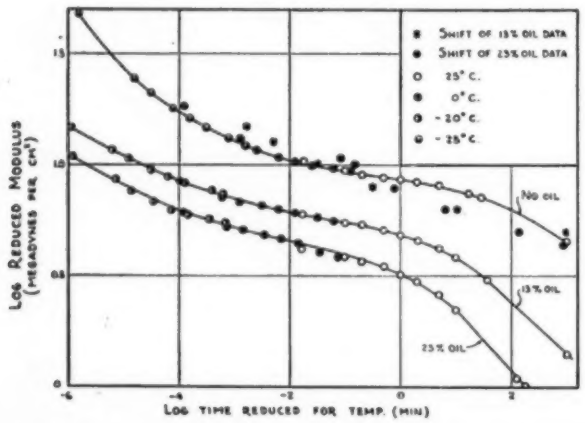


FIG. 4. Creep curves: Polyisobutylene:mineral oil system.

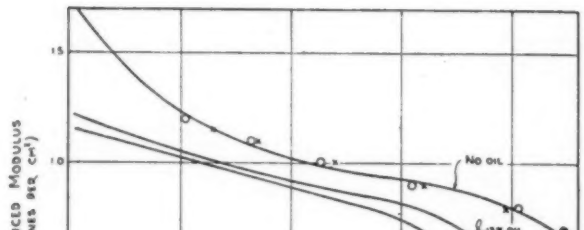

No Oil
13% Oil
23% Oil
Shift of 13% Oil Data
Shift of 23% Oil Data
Log Reduced Modulus (Megadynes per cm²)
Log Time Reduced for Temp. (min)

FIG. 5. Creep curves for polyisobutylene: mineral oil corrected for cross-link dilution.

divided by the gas constant.[6] All values are within experimental error and it is apparent that a change of 1.4 kcal would not be detected. Therefore, for simplicity, it will not be inconsistent to assume that changes in free energy of activation are a result only of changes in "activation energy."

The second system studied was composed of a high molecular weight polyisobutylene plasticized with mineral oil. Temperature-reduced data for the pure polymer and two samples containing the oil are pre-

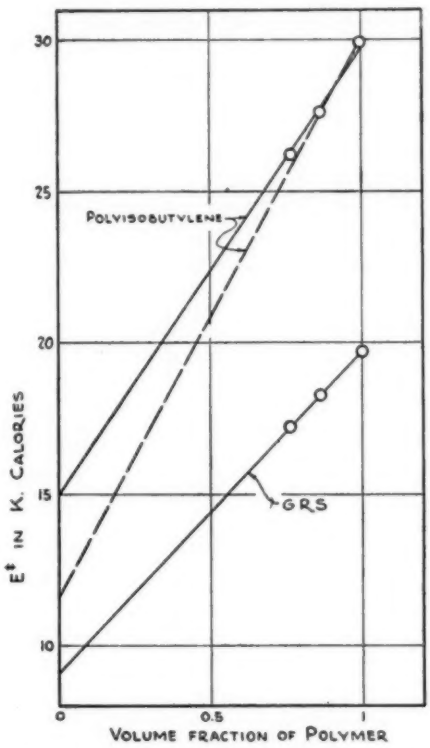


FIG. 6. Energy of activation *vs* composition of plasticized polymers.

[6] Andrews, Hofman-Bang, and Tobolsky, J. Polymer Sci. 3, 669 (1948).

sented in Fig. 4. Unlike the GR-S, this material exhibits a pseudoequilibrium modulus in the region 0.01 to 1 min. Presumably this is a result of chain entanglements[7] and corresponds to region *B* in Fig. 1. Note that the curves for plasticized polymer do not superpose on the unplasticized polymer curve by mere shifts along the log time axis.

Since the number of chain entanglements will vary as the square of the volume fraction v of polymer, multiplying the equilibrium modulus by the factor $1/v^2$ should permit proper shift of the curves of the plasticized material. It would appear that this correction could be properly applied not only to the equilibrium modulus but also to all values of the modulus at longer times. Being a multiplying factor, it will diminish toward zero in the proper fashion as the modulus approaches zero. However, at shorter times than those corresponding to the equilibrium modulus such a correction cannot be the proper one since it should also approach zero at infinitely short times. Furthermore, at the midpoint of the dispersion region, the modulus should be raised about half of the amount that the equilibrium modulus has been changed. Thus an additive term which gradually decreases at shorter times is the most logical correction for all modulus values above the equilibrium modulus. An approximation for this correction is a constant additive term equal to that calculated for the equilibrium modulus.

TABLE II. Plasticizer shift factors and energy of activation temperature 25°C.

GR-S:mineral oil			Polyisobutylene:mineral oil		
Composition (wgt) polymer: plasticizer	Shift factor	$E\ddagger$, kcal	Composition (wgt) polymer: plasticizer	Shift factor	$E\ddagger$, kcal
100:0	0	19.7	100:0	0	30.0
86:13	1.05	18.3	86:13	1.85	27.5
76:23	1.77	17.3	76:23	2.55	26.5

When the plasticized polyisobutylene curves are corrected for the decrease in the number of chain entanglements per gram of polymer in the above fashion, good shifting of the curves is possible. Figure 5 depicts the corrected curves and the excellent shift which results.

Since increased rates of relaxation for the plasticized polymers are attributed to a reduction in the energy of activation for viscoelastic flow, plots of $E\ddagger$ for the two polymers have been prepared. Data are given in Table II and plots in Fig. 6. The points for the GR-S system fall on a good line which extrapolates to 9 kcal as the "energy of activation" for mineral oil. The "activation energy" for mineral oil as determined by the temperature coefficient of viscosity is found to be 10 kcal. For polyisobutylene, the points do not give as good a line, and the extrapolated value is 15 kcal. If the point corresponding to the 23-percent oil composition is ignored, a line through the remaining two points extrapolates to 12 kcal. It is apparent that more extended data are needed to complete the picture.

Plasticizer addition has shown the expected effect of lowering the "equilibrium" value of the modulus and, at the same time, increasing the rate of the relaxation (or retardation) processes. Further dilution with oil would be expected eventually to eliminate region *B* so that processes 1 and 2 would merge. In such a region, addition of plasticizer should give curves which could be shifted along the log time axis to coincide with each other but not with the curve for the pure polymer.

Once again, the apparent activation energy, as measured by the modulus temperature coefficient, shows no significant change with plasticizer addition. The value (30 kcal) however, is about 10 kcal higher than that reported earlier for a lower molecular weight polyisobutylene. No explanation is immediately obvious.

It is apparent that more extended data are needed, so these conclusions should be regarded as tentative.

ACKNOWLEDGMENT

The advice and helpful criticisms of C. A. Dahlquist and other members of the Central Research Department of Minnesota Mining and Manufacturing Company are gratefully acknowledged. The authors also wish to express their appreciation to Professor J. D. Ferry of the University of Wisconsin for valuable discussions concerning this work.

Elasticity measurements were made by W. E. Luedke and K. E. Peterson.

[7] H. Mark and A. V. Tobolsky, *Physical Chemistry of High Polymeric Systems* (Interscience Publishers, Inc., New York, 1950).

Some Rheological Properties of Concentrated Polyisobutylene Solutions*

F. J. PADDEN AND T. W. DEWITT
Mellon Institute, Pittsburgh 13, Pennsylvania

(Received April 12, 1954)

Three types of measurements have been made on concentrated polyisobutylene solutions. Dynamic measurements have been extended to low frequencies and the dynamic viscosities and rigidities determined in the region close to that in which pure viscous flow predominates. A modified Stormer viscometer has been used to determine the apparent (non-Newtonian) viscosities of the same solutions. Pressures developed (normal stress) at the inner cylinder have also been measured in a concentric cylinder apparatus with the inner cylinder rotating. It is noted that superposition procedures currently used for dynamic data are apparently applicable to the viscosity and pressure data with the rate of shear taking the place of frequency. Relations between these phenomena are discussed in the light of a theory proposed by one of the authors.

I. INTRODUCTION

CONCENTRATED polymer solutions belong to a class of materials, frequently called viscoelastic, that show several types of anomalous behavior. One of the best known of these is their response to repetitive stresses; they behave partly as solids and partly as liquids and their responses may be analyzed in terms of both a modulus and a viscosity.[1] They also frequently show what is known as "non-Newtonian" viscosity,[2] which commonly designates an inverse dependence of the apparent viscosity coefficient on the rate of shear. Although this type of behavior is not usually treated as being associated with the above-mentioned dynamic behavior, it is certainly a prevalent feeling among workers in the field that there should be a connection. In this same category is the so-called Weissenberg,[3] or "normal stress" effect[4–6] wherein radial pressure gradients appear when the material is subject to shear either between rotating parallel plate or concentric cylinders.

Recently one of the authors has proposed[5] an equation of state for such materials which relates the three aforementioned effects and assigns them a common origin.

The proposed equation of state is:

$$p_{ij}+\tau(\dot{p}_{ij}-\omega_{i\alpha}p_j{}^{\alpha}-\omega_{j\alpha}p_i{}^{\alpha})=2\eta d_{ij}, \tag{1}$$

where p_{ij} (also $p_i{}^{\alpha}$ and $p_j{}^{\alpha}$) is a stress component; $\dot{p}_{ij}$, its total time derivative; $\omega_{i\alpha}$ (also $\omega_{j\alpha}$), a component of the vorticity (tensor) of the fluid motion; τ, a relaxation time equal to η/G, where η is the Newtonian viscosity coefficient and G is a shear modulus; and d_{ij}, a component of the rate of deformation. When τ is sufficiently small, Eq. (1) reduces to the usual Newtonian viscosity expression. If the cross terms $\omega_{i\alpha}p_j{}^{\alpha}$ are small compared with $\dot{p}_{ij}$, it reduces to the equation for a Maxwell element. In this case, if $\dot{p}_{ij}$ is given a sinusoidal time dependence, the usual formulation,[1] in terms of a complex viscosity or modulus, becomes applicable and one finds

$$G'=G\tau^2\omega^2/(1+\tau^2\omega^2), \tag{2}$$

and

$$\eta'=\eta/(1+\tau^2\omega^2), \tag{3}$$

where G' and η' are the dynamic modulus and dynamic viscosity, respectively; G is the modulus of the Maxwell element; η is the Newtonian viscosity coefficient; τ is the relaxation time; and ω is the angular frequency. If the cross terms are not small compared to $\dot{p}_{ij}$, the dynamic case would be nonlinear. In a steady-flow experiment, $\partial p_{ij}/\partial\tau$ is zero, and if τ is not small, the cross terms are important and the steady-state case is nonlinear. Reduction of Eq. (1) to cylindrical geometry and application of the resulting expressions to the case of rotating concentric cylinders, yields

$$\langle r\theta\rangle_{Ni}=\frac{\eta[r(\partial\omega/\partial r)]}{1+\tau^2[r(\partial\omega/\partial r)]^2},$$

where $\langle r\theta\rangle_{Ni}$ is the stress in the angular direction; r is the radius; ω is the angular velocity; and $r(\partial\omega/\partial r)$ is the rate of shear. This is the same as that for a rotating cylinder viscometer if

$$\eta_a=\frac{\eta}{1+\tau^2[r(\partial\omega/\partial r)]^2} \tag{4}$$

is taken as the "apparent" viscosity coefficient usually calculated for non-Newtonian behavior. The cross terms also lead to the prediction that there should be a pressure developed at the inner cylinder given by

$$\Delta p=P=\frac{\eta\tau[r(\partial\omega/\partial r)]^2}{1+\tau^2[r(\partial\omega/\partial r)]^2}$$
$$-\frac{G}{2}\ln\left[\tau^2\left(r\frac{\partial\omega}{\partial r}\right)^2+1\right]+C, \tag{5}$$

* The work discussed herein was performed as part of the research project sponsored by the Reconstruction Finance Corporation, Office of Synthetic Rubber, in connection with the Government Synthetic Rubber Program.

[1] Ferry, Sawyer, and Ashworth, J. Polymer Sci. 2, 593 (1947).
[2] M. Reiner, *Deformation and Flow* (H. K. Lewis and Company, London, 1949).
[3] K. Weissenberg, Nature 159, 310 (1947); Proc. Intern. Congr. Rheol. I, 29, 46 (1948).
[4] R. S. Rivlin, Nature 161, 567 (1948); Proc. Roy. Soc. (London) A193, 260 (1948).
[5] T. W. DeWitt (to be published).
[6] H. W. Greensmith and R. S. Rivlin, Trans. Roy. Soc. (London) A245, 399 (1953).

where Δp is the pressure rise measured from that at zero rate of shear as a reference. In the case of rotating parallel plates (an experimental arrangement of this kind has been used by Garner, Nissan, and Wood,[7] and Greensmith and Rivlin[6]) a pressure gradient in the radial direction is also predicted, and the dependence of the pressure observed at the surface of the upper plate is given by:

$$\Delta p = p_0 - P = \frac{-\eta\tau r^2(\partial\omega/\partial z)^2}{1+\tau^2 r^2(\partial\omega/\partial z)^2} + \frac{G}{2}\ln\left[r^2\tau^2\left(\frac{\partial\omega}{\partial z}\right)^2+1\right], \quad (6)$$

where Δp is measured from the center of the plates ($r=0$) and $\partial\omega/\partial z$, the angular velocity gradient, is independent of r and z.

Equation (1) and its consequences, Eqs. (2)–(6), form the basis for correlating several different types of measurements on the mechanical behavior of concentrated polymer solutions.

A considerable amount of data on the mechanical behavior of concentrated solutions of polyisobutylene have been accumulated in this laboratory over the past several years. These have been collected and, together with some supplementary measurements, are reported here and used as a test of some of the consequences of the proposed equation of state.

II. MATERIALS AND METHODS

The bulk of the data reported were obtained on solutions of polyisobutylene (Vistanex B-100 of viscosity average molecular weight $\sim 10^6$) in decalin. The primary exception is the pressure rise data in the parallel plate apparatus which were obtained on solutions of Vistanex B-80 and B-140 in tetralin. Zero shear viscosities on all solutions except those used in the parallel plate apparatus were obtained by the falling ball method. In each case the Faxen correction was applied and usually at least two balls of different size were used in order to insure that effectively zero rates of shear had been obtained. Viscosities for the solutions of B-80 and B-140 in tetralin were originally obtained with a Brookfield viscometer and empirically extrapolated to zero rate of shear. The results of this unsatisfactory method are highly suspect and cannot be taken too seriously. Since the original solutions are no longer available, and the results are probably at least of the correct order of magnitude, they have been used, with some misgivings, in order to make use of the pressure rise data.

Dynamic measurements in the frequency range of 10–100 cycles per sec were made with a mechanically driven coaxial cylinder apparatus described elsewhere.[8] These data were supplemented by low-frequency measurements in an electrically driven torsion pendulum apparatus.[9] This instrument has previously been used for solid polymer samples and was modified for use with liquids by equipping it with a cup and cylindrical bob so that it employed essentially the same geometry as the higher frequency apparatus. With this instrument, measurements were made down to about 0.005 cycle per sec.

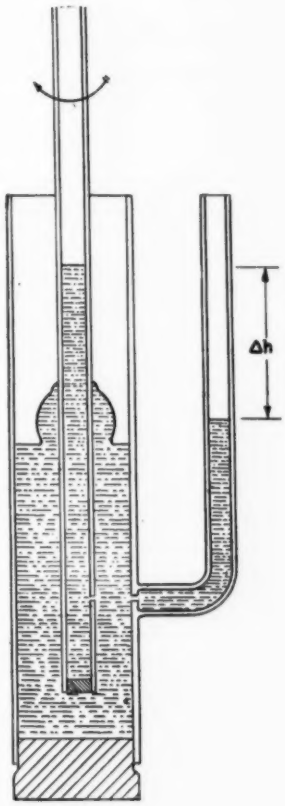


FIG. 1. Coaxial cylinder apparatus for normal stress effect measurements.

Steady-flow viscosities were measured in a modified Stormer viscometer. The liquid sample occupied the annular space between a fixed cylindrical cup and a rotating coaxial cylindrical bob. The bob was rotated by the usual gear, pulley, and weight arrangement. The instrument was calibrated by plotting a family of curves of rate of rotation *vs* viscosity—one for each weight. A series of low molecular weight Oronite polyisobutylene oils[10] and a sample of "Abitol"[11] were used as calibrating liquids. These liquids were shown to be Newtonian for the rates of shear used and covered a range of 20 to 1500 poises. The cup containing the sample was immersed in a water bath controlled to ± 0.02°C and the

[7] Garner, Nissan, and Wood, Trans. Roy. Soc. (London) **A243**, 37 (1950).

[8] Markovitz, Yavorsky, Harper, Zapas, and DeWitt, Rev. Sci. Instr. **23**, 430 (1952).

[9] T. E. Morrisson and T. W. DeWitt, Phys. Rev. **85**, 708 (1952).

[10] Supplied by the Oronite Chemical Company, San Francisco, California.

[11] Supplied by the Hercules Powder Company, Wilmington, Delaware.

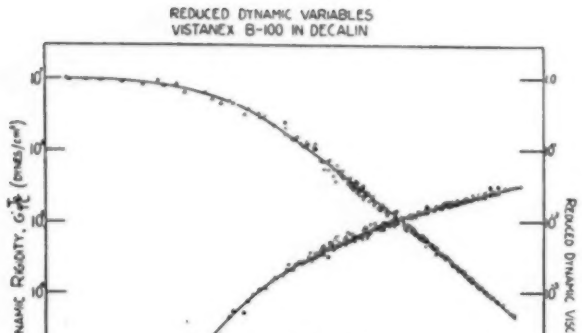


FIG. 2. Reduced dynamic rigidity and reduced dynamic viscosity as a function of reduced frequency for solutions of polyisobutylene in decalin.

whole apparatus was housed, in an insulated box controlled to ±0.2°C. All measurements were made at 35°C.

Pressures developed when these solutions are sheared were observed in the simple apparatus shown in Fig. 1. A length of hollow glass tubing is centered by means of metal spacers in a test tube. The center tube, which was rotated at speeds from 0 to 15 rps, contains a small hole near the bottom. Opposite it is another hole in the wall of the test tube leading to a vertical tube of the same diameter as the inner tube. The heights to which liquid rises in the rotating tube and the side tube measure the pressures at the inner and outer walls, respectively.

Data from the rotating parallel plate instrument were obtained several years ago, and are, for a number of reasons, somewhat questionable. The apparatus is essentially the same as that described by Greensmith and Rivlin,[8] although mechanically much cruder. As a matter of fact, it was originally constructed by Rivlin for some of his preliminary work on this effect. It consists of a circular horizontal bottom plate with low vertical sides. This large diameter cup holds the liquid sample, and is rotated by a motor and gear box. Parallel to and above this, at a distance of about a centimeter or less, is a stationary circular plate that carries several radially placed vertical glass tubes that have access to the liquid sample through small holes. When the lower plate is rotated, the liquid, if it shows the Weissenberg effect, rises in the vertical tubes and the height to which it rises is taken as a measure of the pressure developed at the surface of the stationary plate.

III. RESULTS

The dynamic results are summarized in Fig. 2, where the data are plotted in terms of the reduced variables introduced by Ferry.[12,13] Reduced variables are based on Eqs. (2) and (3), which, assuming a distribution of relaxation times, and superposition of the effects due to different relaxation times, can be written:

$$G'=\sum_i \frac{G_i\tau_i^2\omega^2}{1+\tau_i^2\omega^2}, \tag{7}$$

and

$$\eta'=\sum_i \frac{\eta_i}{1+\tau_i^2\omega^2}=\sum_i \frac{G_i\tau_i}{1+\tau_i^2\omega^2}, \tag{8}$$

where G_i is the modulus associated with the ith relaxation time τ_i, and η, the zero shear viscosity, is $\sum_i G_i\tau_i$. It is then assumed that all G_i depend identically on, and are proportional to, absolute temperatures and concentration, and one has $G_i=G_{i0}(Tc/T_0c_0)$, where G_{i0} is the value of G_i at the reference temperature and concentration of T_0 and c_0. It is further assumed that all τ_i also depend identically on temperature and concentration and that there are proportionality constants a_T and a_c such that $\tau_i=a_Ta_c\tau_{i0}$ where τ_{i0} is the ith relaxation time in the reference state. When these assumptions are incorporated into the expressions for the dynamic rigidity and viscosity, it appears that if a reduced dynamic modulus is defined as $G_R'=G'(T_0c_0/Tc)$, a reduced dynamic viscosity as η'/η, and a reduced frequency as $a_Ta_c\omega$, data taken under a variety of conditions should form a single curve, which gives the values of the quantities of interest for the arbitrarily chosen reference state. With the observation that, since,

$$\eta=\sum G_i\tau_i=\frac{Tc}{T_0c_0}a_Ta_c\sum G_{i0}\tau_{i0}=\frac{Tc}{T_0c_0}a_Ta_c\eta_0$$

$$a_Ta_c=(\eta T_0c_0/\eta_0Tc),$$

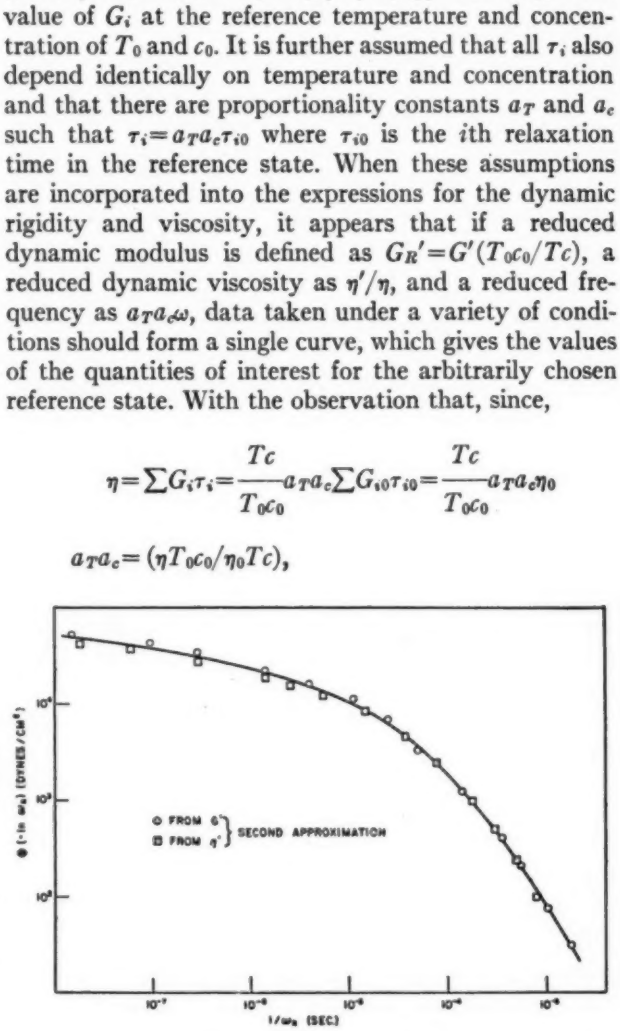


FIG. 3. Distribution of relaxation strengths for solutions of polyisobutylene in decalin, calculated using the second approximation of Ferry (see reference 13).

[12] J. D. Ferry, J. Am. Chem. Soc. 72, 3746 (1950).
[13] Ferry, Fitzgerald, Grandine, and Williams, Ind. Eng. Chem. 44, 703 (1952).

and with the choice of a standard state such that c_0 and η_0 are unity, the reduced variables plotted in Fig. 1 are obtained. The density of points precludes individual identification of the concentrations, temperatures, and frequencies but the data include concentrations from 5 to 22 percent, temperatures from 5°C to 35°C, and frequencies from 0.005 to 100 cycles per second. Solid square points are taken from data reported by Ferry on solutions of polyisobutylene of about the same molecular weight in xylene. Ferry's data were obtained by wave propagation and transducer methods, both of which are completely distinct from the techniques used in the present work. Points designated as obtained with the torsion pendulum are new data that complete the low-frequency end of the spectrum for this system to the point where rigidity has virtually disappeared and the dynamic viscosity is equal to the zero shear steady-state viscosity.

When it is assumed, as is generally necessary, that a system requires a continuous distribution of relaxation times for its description, this distribution is commonly given in terms of a function ϕ, defined so that $\phi d\ln\tau$ represents the contribution to the rigidity from relaxation times between $\ln\tau$ and $\ln\tau+d\ln\tau$. This may be approximated by graphical methods from the data in Fig. 2. The method of Ferry,[13] including the second approximation, has been applied to these data and the results are given in Fig. 3.

Some typical results from the Störmer measurements are shown in Fig. 4. It is a consequence of the similarity between Eqs. (3) and (4) that if a reduced shear rate $[\tau(\partial\omega/\partial\tau)]$ and a reduced viscosity (η_a/η) is defined in the same way as was done for the dynamic case, an entirely analogous superposition of data should be possible. A partial test of this is shown in Fig. 5 where the use of reduced variables is seen to bring data at several concentrations into a single curve. It is a further consequence of the form of Eq. (4) that if the rate of shear, $\tau(\partial\omega/\partial\tau)$, is identified with frequency ω, the dynamic viscosity η' should be the same function of ω as η_a is of $\tau(\partial\omega/\partial\tau)$. In the case of the coaxial cylinder

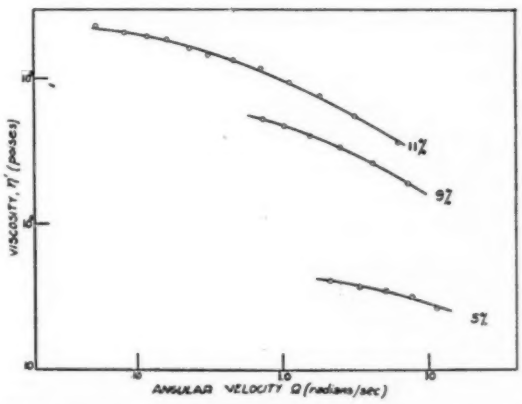


FIG. 4. Apparent steady-state viscosity of several concentrations of polysiobutylene in decalin as a function of the angular velocity of the inner cylinder of a coaxial cylinder viscometer.

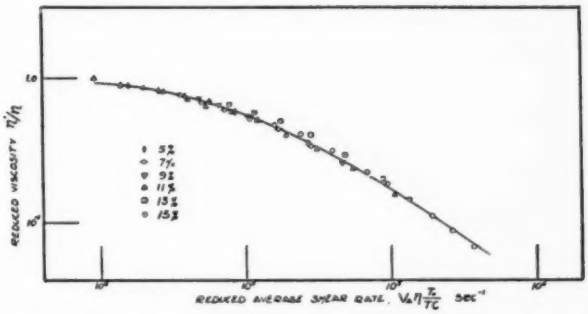


FIG. 5. Reduced steady-state viscosity from coaxial cylinder viscometer measurements plotted as a function of reduced average shear rate for various concentrations of polyisobutylene in decalin.

instrument, the frequency, in a dynamic experiment, is constant across the gap, whereas the rate of shear in a steady-flow experiment is not. It is therefore expected that an average shear rate would have to be used in any quantitative comparison between Eqs. (3) and (4). In computing the reduced rate of shear used in Figs. 5 and 6, an average suggested by Andrade[14] has been used. This average is calculated from the expression

$$V_A=\frac{\int_{R_i}^{R_0} 2\pi r\cdot r\frac{\partial\omega}{\partial r}dr}{\pi(R_i^2-R_0^2)}=\frac{4R_i^2R_0^2\Omega}{(R_i^2-R_0^2)^2}\ln\frac{R_0}{R_i}, \qquad (9)$$

where R_i is the radius of the inner cylinder, R_0 is that of the outer, and Ω is the angular velocity of the inner cylinder. In calculating this average rate of shear, an approximation is involved in that it has been assumed that the rate of shear is the same as that for the Newtonian case. In Fig. 6, the data in Fig. 4 are plotted along with those for the dynamic viscosity. It is seen that when the data are plotted as reduced variables and the average rate of shear or the frequency alternately used for the steady state and dynamic data, respectively, all the data can be represented by a single curve.

If the same reduced variable treatment is applied to Eq. (5), one should be able to plot a reduced pressure rise variable, $2\Delta p(T_0/Tc)$, against the reduced rate of shear, $V(T_0\eta/Tc)$, and obtain a single curve for different concentrations. Figure 7 shows the pressure rise in the inner tube of the concentric cylinder apparatus plotted against the angular velocity of the inner tube. Figure 8 shows the same data plotted against the reduced rate of shear at the inner wall, calculated from the Newtonian expression. These results demonstrate the applicability of reduced variables to this phenomenon as predicted by the theory. In the case of the parallel plate experiments, the form of Eq. (6) is exactly the same as that of Eq. (5) and therefore pressure differences measured radially from the center of the plate should be directly comparable to those from the

[14] E. N. da C. Andrade, *Viscosity and Plasticity* (W. Heffer and Sons, Ltd., Cambridge, 1947), p. 40.

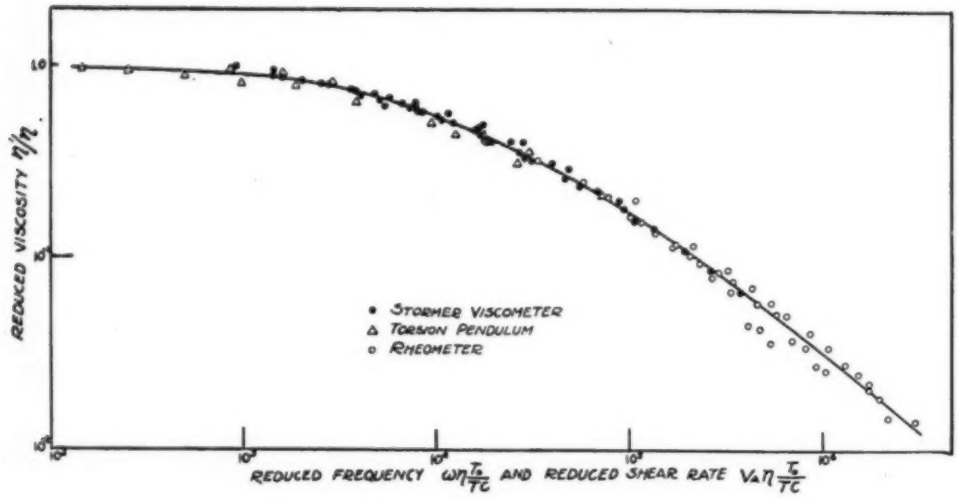


FIG. 6. Superposition of reduced dynamic and steady-state viscosities for solutions of polyisobutylene in decalin.

concentric cylinder apparatus, provided $r(\partial\omega/\partial z)$ is taken as the rate of shear and reduced variables are used. A test of the superposition procedures for this case is shown in Fig. 9. In the same figure, the concentric cylinder data from Fig. 8 and the dynamic modulus data from Fig. 2 are reproduced for comparison. An additional factor has been included in the reduced variables in Fig. 9 to take into account molecular weight differences as previously found necessary for dynamic data.[15] Since the original samples were no longer available, these factors were obtained from dynamic measurements on solutions of polyisobutylenes of the same nominal molecular range. The data of Greensmith,[16] which are included in Fig. 9, were also corrected by an estimated molecular weight factor and this, together with his extrapolated value for the zero shear viscosity of his solution, was used in calculating a reduced shear rate.

The parallel plate data, in Fig. 9, superposes about as well as might be expected when the uncertainties of the data are taken into account. Further than this, however, the quantitative, and even the qualitative, predictions of Eq. (6) are entirely inadequate. Aside from its quantitative failure, this expression would lead one to predict that the observed pressure would increase as one proceeds from the center of the apparatus to the rim, and this, of course, is just the reverse of what is observed. In contrast to this, Eq. (5) reasonably well predicts the observed behavior in the concentric cylinder apparatus. Indeed, if the admittedly somewhat dubious, but for the present purposes useful, approximation is made of using the first term in the expansion of the logarithm in Eq. (5), it becomes

$$P=\frac{G\tau^2[r(\partial\omega/\partial r)]^2}{1+\tau^2[r(\partial\omega/\partial r)]^2}-\tfrac{1}{2}G\tau^2\left(r\frac{\partial\omega}{\partial r}\right)^2. \qquad (10)$$

For rates of shear sufficiently small so that the denominator of the first term may not differ too greatly from unity, this can be written, after summing over all relaxation times,

$$P\cong G'/2,$$

where G' is the dynamic modulus for a reduced frequency equal to the reduced shear rate. It is seen in Fig. 9 that the observed pressure rises are, in this case, at least of the correct order of magnitude.

If the same treatment is applied to the cone and plate type of apparatus, where the rate of shear remains sensibly constant over the radius of the apparatus, one would have $dp/d\ln r=-G\tau^2K^2/(1+\tau^2K^2)$ where K is a constant involving the angular rate of shear and the angle between the cone and plate but independent of the radius. Integrating between r and R, the radius of the apparatus, one has

$$p=\frac{-G\tau^2K^2}{1+\tau^2K^2}\ln\frac{r}{R}, \qquad (11)$$

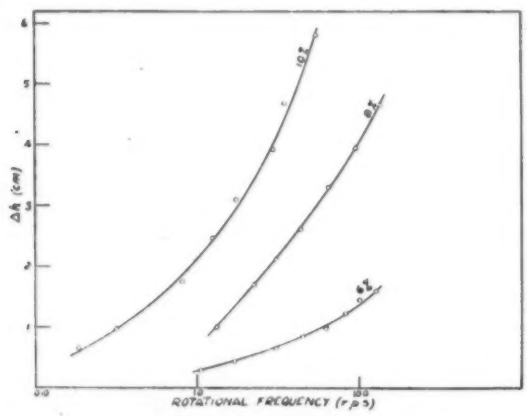


FIG. 7. Pressure rise in the coaxial cylinder instrument as a function of the rotational speed of the inner cylinder for three solutions of polyisobutylene in decalin.

[15] Padden, Markovitz, and DeWitt, Phys. Rev. 91, 217 (1953).
[16] H. W. Greensmith, Ph.D. thesis, University of London, 1952.

and

$$P=\frac{G\tau^2K^2}{1+\tau^2K^2}+\frac{G\tau^2K^2}{1+\tau^2K^2}\ln\frac{R}{r}. \qquad (12)$$

In this case, the pressure is predicted to rise logarithmically with r as the center is approached. This behavior appears to check with that observed by Roberts[17] for apparatus of this type.

Thus the proposed treatment seems to account, at least semiquantitatively for the viscosity behavior of viscoelastic systems, and to correctly predict the presence of associated pressure effects. It adequately predicts the pressure behavior in the concentric cylinder apparatus and is at least partially in agreement with observed effects in the cone and plate apparatus. In the parallel plate type of experiment, it predicts results qualitatively the reverse of those observed. Whether this failure is to be attributed to a fundamental inadequacy of the theory or to improper application of the theory to this particular case remains to be examined in future work. Further than this, it is to be cautioned

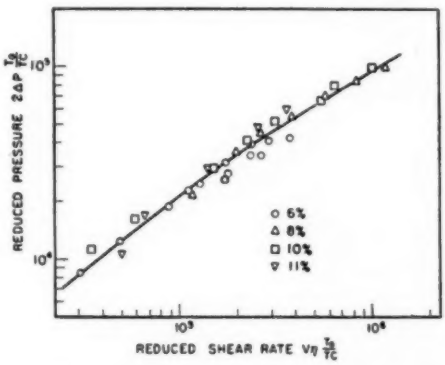


FIG. 8. Reduced variables used to obtain a single curve for pressure rise as a function of reduced rate of shear at the inner cylinder for coaxial cylinder data on four concentrations of polyisobutylene in decalin.

that the present data, assembled as they have been from a variety of experiments, many of which were not undertaken with the present analysis in mind, are but preliminary tests of the proposed correlation. Much more extensive data will be required before they can be considered anything but tentative. It must also be remembered that in the application of Eq. (1) to the data in this paper, it has been assumed that rates of shear do not deviate too severely from the Newtonian case and that it is therefore permissible to calculate rates of shear or average rates of shear from the Newtonian

[17] J. E. Roberts, Preprints of papers read at Second International Congress on Rheology, Butterworths, London, 1953, p. AA161.

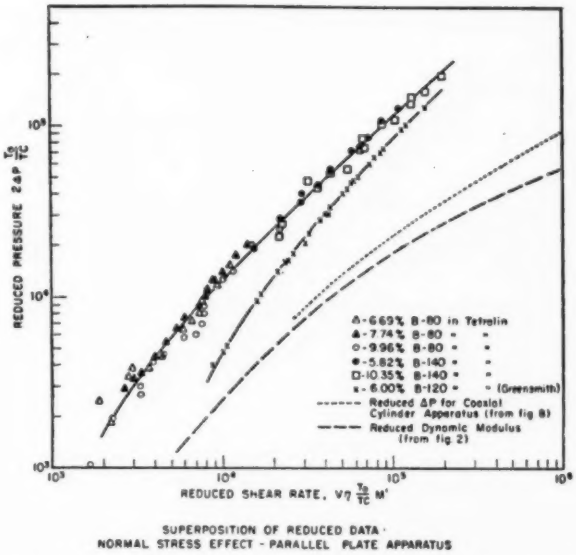

REDUCED PRESSURE 2ΔP τ0/τC
REDUCED SHEAR RATE, Vη τ0/τC M'
Δ-6.69% B-80 in Tetralin
▲-7.74% B-80 " "
○-9.96% B-80 " "
●-5.82% B-140 " "
□-10.35% B-140 " "
×-6.00% B-120 " " (Greensmith)
Reduced ΔP for Coaxial Cylinder Apparatus (from fig 8)
Reduced Dynamic Modulus (from fig 2)
SUPERPOSITION OF REDUCED DATA
NORMAL STRESS EFFECT - PARALLEL PLATE APPARATUS

FIG. 9. Reduced pressure rise as a function of reduced rate of shear for parallel plate normal stress measurements on solutions of polyisobutylenes in tetralin. A reduced curve plotted from the data of Greensmith (reference 16) and the curves from Figs. 2 and 8 are included for comparison.

expressions. To what extent this is a justified assumption is not known. There is also something arbitrary about the choice of the average rate of shear. The Andrade average was used for the viscosity data because it gave good agreement with the dynamic data but this may be fortuitous. This average cannot be generally applicable since it displays the annoying feature of going to zero in the limit of large outer cylinders and this doesn't seem physically reasonable. Thus the question of how best to handle cases of nonuniform shear without full integration of the basic equations must for the moment remain open. Data obtained with a wider range of geometries may provide a clue as to the best approach. In view of the complications that seem to be introduced in the interpretation of the experiments involving nonuniform rates of shear, it is important that more experimental work be undertaken, both for viscosity and normal stress measurements, in the cone and plate type of apparatus, where the rate of shear is uniform over the apparatus.

IV. ACKNOWLEDGMENT

We are indebted to Miss Alba Anania and Mr. Dean Sensenbaugh for having taken some of the early data on the parallel plate apparatus, and also to Mr. Louis Zapas for obtaining the low-frequency torsion pendulum data. Mr. Rivlin kindly made Greensmith's thesis available to us.

Elastoviscous Properties of Amorphous Polymers in the Transition Region. II. Stress-Relaxation of Butadiene-Styrene Copolymers

E. CATSIFF AND A. V. TOBOLSKY
Frick Chemical Laboratories, Princeton University, Princeton, New Jersey

(Received April 12, 1954)

Experimental data have been obtained on the stress-relaxation modulus $E_r(t)$ of several copolymers in their transition regions. The new data are expressed as master curves which can be fitted to the Gauss Error Integral form of reduced equation, as previously proposed. It appears possible to characterize the time and temperature-dependence of E_r of amorphous polymers in general by means of the parameters of this equation: E_1 (glassy-state modulus), E_2 (rubbery-state modulus), T_d (distinctive temperature, related to second-order transition temperature), K_d (characteristic relaxation time at T_d), and h (parameter related to steepness of master curve). In the butadiene-styrene series, the effect of composition on these parameters has been determined.

INTRODUCTION

IN the first paper of this series[1] (hereinafter referred to as paper I), it was shown that the composite stress-relaxation curves of a GR-S gum vulcanizate and of polymethyl methacrylate (PMMA) in their respective transition regions could be adequately reproduced by an equation of the Gauss Error Integral form:

$$\log E_r(t/K)=C-B\,\mathrm{erf}\,(h\log t/K), \qquad (1)$$

where

$E_r(t/K)=E_{r,T}(t)=$stress/strain in a sample maintained at a constant small strain for a time t at temperature T;

$K=$characteristic relaxation time, a function of temperature only for a given polymer;

$$C=\frac{\log E_1+\log E_2}{2};$$

$$B=\frac{\log E_1-\log E_2}{2};$$

$E_1=$quasistatic glassy modulus;

$E_2=$quasistatic rubbery modulus;

$\mathrm{erf}\,x=2\pi^{-\frac{1}{2}}\int_0^x e^{-u^2}du$, the error integral[2]; and

$h=$adjustable parameter of the Gauss error curve.

Furthermore, it was shown that the temperature-dependence of K for both polymers was

$$h\log K_R=f(T_R);$$

where $K_R=K/K_d$; $K_d=K$ at T_d; $T_R=T/T_d$; and $T_d=$distinctive temperature, which is related to second-order transition temperature.

In the range $0.95<T_R<1.05$, which includes nearly all of the transition region, we approximated $f(T_R)$ by $-36(T_R-1)$, although the actual shape of the function was slightly sigmoid, with the inflection point at $T_R=1.00$.

In this paper, the results of stress-relaxation studies on four additional amorphous copolymers are presented. Three of these are gum vulcanizates of butadiene-styrene copolymers containing 40, 50, and 70 wt-percent of styrene, respectively, while the fourth is Paracril 26 gum vulcanizate. With the two polymers described in paper I, they cover a broad range of transition temperatures.

Equation (1) has been fitted to the new data satisfactorily. It has been necessary to recognize that $f(T_R)$ in Eq. (2) is definitely a sigmoid curve; we are not yet able to specify its exact locus, but the experimental points obtained thus far do not show excessive scatter. The general validity of the reduced equation for viscoelastic behavior appears to be strengthened by these results.

EXPERIMENTAL RESULTS

The apparatus used was described in paper I. An improved spring dynamometer, manufactured by the Schaevitz Engineering Corporation, was used to measure stress. Strains were kept below 10 percent to minimize the effects of any nonlinearity in the stress-strain curve.

The Paracril 26 vulcanizate was compounded of 100 parts Paracril 26 (26 percent acrylonitrile, 74 percent butadiene copolymer), 1 part stearic acid, 5 parts zinc oxide, 1 part Santocure, and 1.5 parts sulfur and cured 30 minutes at 310°F.[3] The three butadiene-styrene copolymers were compounded of 100 parts copolymer, 2 parts sulfur, 5 parts zinc oxide, 0.5 parts stearic acid, and 1.5 parts Captax and cured 100 minutes at 270°F.[4]

[1] Bischoff, Catsiff, and Tobolsky, J. Am. Chem. Soc. 74, 3378 (1952).

[2] Tables of the error integral may be found in Jahnke and Emde, *Tables of Functions* (B. G. Teubner, Leipzig and Berlin, 1933); J. W. Mellor, *Higher Mathematics for Students of Chemistry and Physics* (Longmans, Green and Company, London, 1909); and many other places.

[3] The compounding and curing were carried out by the Thiokol Corporation, Trenton, New Jersey.

[4] The compounding and curing were carried out through the courtesy of Mr. Kenneth E. Creed, Jr., of Monsanto Chemical Company on copolymers provided by the Firestone Tire and Rubber Company.

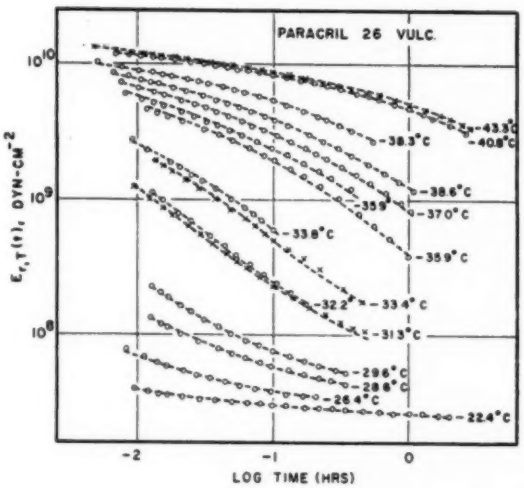


FIG. 1. Stress-relaxation curves for Paracril 26 gum vulcanizate.

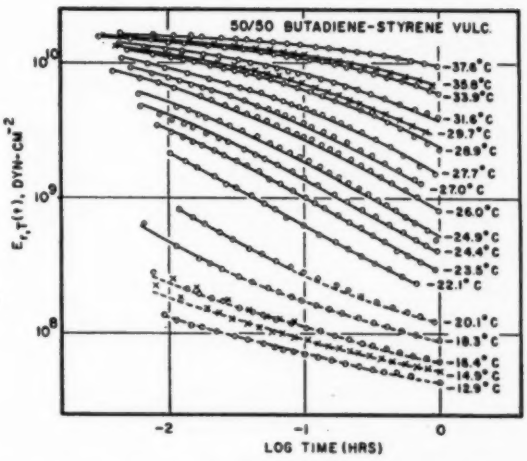


FIG. 3. Stress-relaxation curves for 50/50 butadiene-styrene copolymer gum vulcanizate.

The stress-relaxation curves at various temperatures for the four copolymers are shown in Figs. 1–4. In some cases duplicate points are shown to indicate the reproducibility of the data. The lines are portions of the composite master curve for each polymer which have been shifted along the log-time axis as described in the next section.

The most serious sources of experimental error were insufficiently precise temperature control and insufficiently accurate strain measurement. Several runs were discarded because of temperature drift. Any drift greater than 0.2°C was considered serious enough to invalidate a run. Strain measurement was complicated by imperfect alignment of the clamps, so that some shear strain was involved along with the pure tensile strain on which the computations were based.

This effect was most noticeable when the sample length was changed. For the 50/50 copolymer, it proved necessary to use a small arbitrary strain correction factor to bring the data at the four lowest temperatures (which were obtained on a long sample) into agreement with the remaining data. In consequence, it was found best to confine the study of the 30/70 copolymer to a single sample length. The absolute values of the moduli given may be incorrect, but the relative values should be quite accurate.

CONSTRUCTION OF MASTER CURVES AND FITTING OF GAUSS ERROR INTEGRAL

As can be seen from Figs. 1–4, the shape of the stress-relaxation curve of a given polymer is virtually unaffected by temperature, but the curve is displaced horizontally by change of temperature. The method of combining these displaced curves to form a single master curve was described in paper I. This master curve is conveniently plotted as $\log E_r(t/K)$ *versus* $\log t/K$, as is done for the four copolymers in Figs. 5–8. Table I gives the K values used to construct the master curves. The irregularities and order-reversals found in

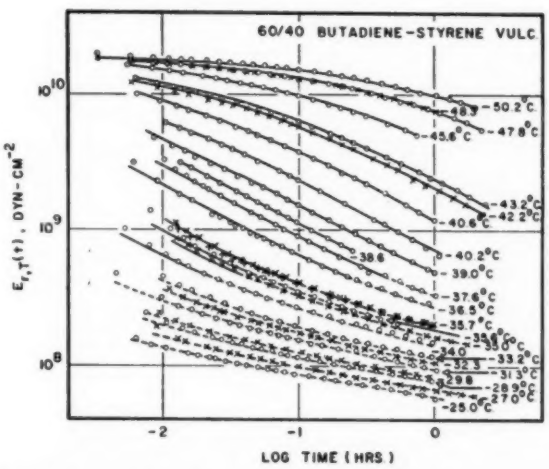


FIG. 2. Stress-relaxation curves for 60/40 butadiene-styrene copolymer gum vulcanizate.

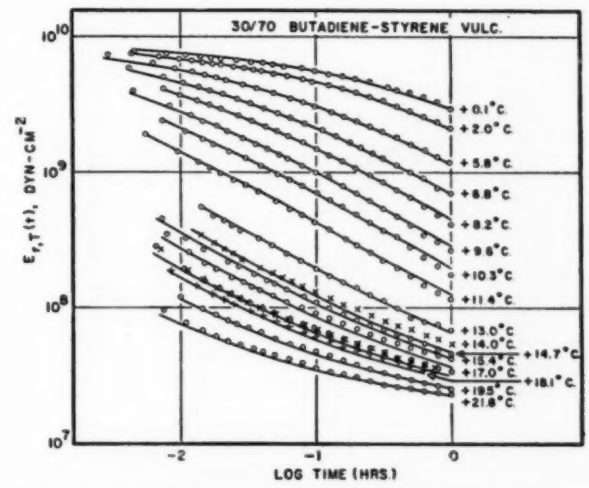


FIG. 4. Stress-relaxation curves for 30/70 butadiene-styrene copolymer gum vulcanizate.

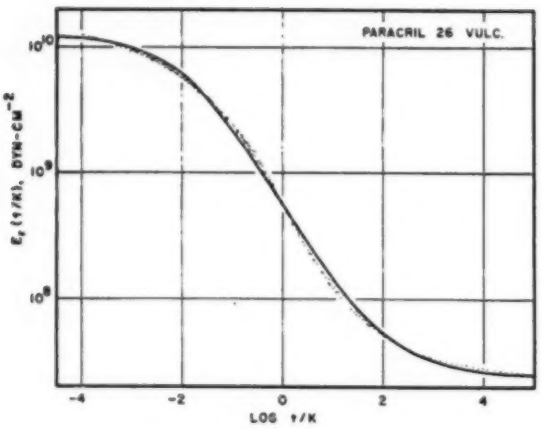


FIG. 5. Master stress-relaxation curve for Paracril 26 gum vulcanizate.

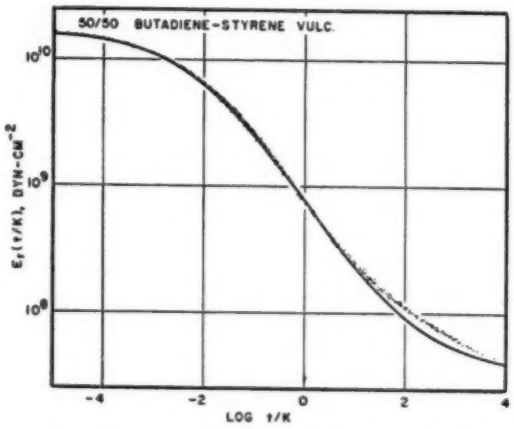


FIG. 7. Master stress-relaxation curve for 50/50 butadiene-styrene copolymer gum vulcanizate.

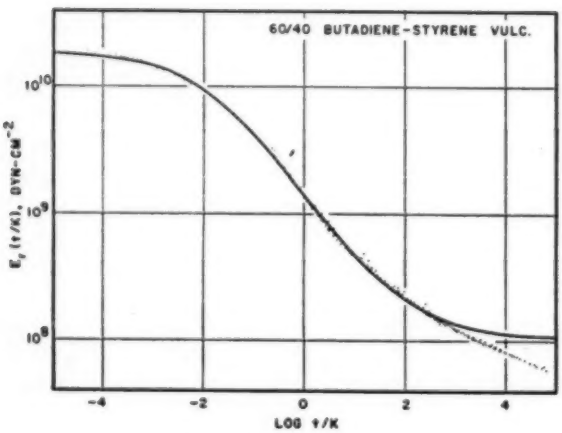


FIG. 6. Master stress-relaxation curve for 60/40 butadiene-styrene copolymer gum vulcanizate.

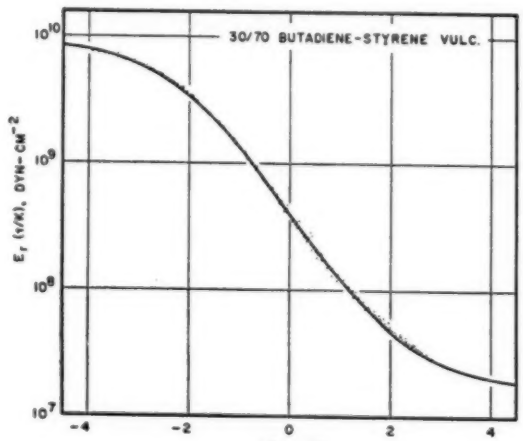


FIG. 8. Master stress-relaxation curve for 30/70 butadiene-styrene copolymer gum vulcanizate.

TABLE I. Temperature dependence of characteristic relaxation time K in hours.

Paracril 26		60/40 Butadiene-Styrene		50/50 Butadiene-Styrene		30/70 Butadiene-Styrene	
Temp. °C	log$K$	Temp. °C	log$K$	Temp. °C	log$K$	Temp. °C	log$K$
−43.3	1.81	−50.2	2.07	−37.6	2.60	0.1	1.83
−40.8	1.69	−48.3	1.67	−35.8	2.06	2.0	1.45
−38.6	0.48	−47.8	1.67	−33.9	1.91	5.8	0.87
−38.3	0.88	−45.6	1.07	−31.6	1.38	6.8	0.45
−37.0	0.22	−43.2	0.45	−29.7	1.11	8.2	0.11
−35.9	−0.01	−42.2	0.33	−28.9	0.96	9.6	−0.25
	−0.19	−40.6	−0.11	−27.7	0.56	10.3	−0.54
−33.8	−0.96	−40.2	−0.56	−27.0	0.28	11.4	−0.94
−33.5	−1.06	−39.0	−0.92	−26.0	0.06	13.0	−1.57
−32.2	−1.48	−38.6	−1.19	−24.9	−0.28	14.0	−1.97
−31.3	−1.53	−37.6	−1.37	−24.4	−0.52	14.7	−2.09
−29.6	−2.46	−36.5	−1.64	−23.5	−0.78	15.4	−2.30
−28.6	−2.82	−35.7	−2.14	−22.1	−1.16	17.0	−2.59
−26.4	−3.60	−35.6	−2.30	−20.1	−1.88	17.1	−2.60
−22.4	−4.54	−35.0	−2.14	−18.3	−2.38	18.1	−2.71
			−2.40	−16.4	−3.02	19.5	−3.03
		−34.0	−2.68	−15.8	−2.96	21.8	−3.43
		−33.2	−3.14	−14.9	−3.32		
		−32.3	−3.30	−12.9	−3.76		
		−31.3	−3.52				
		−29.8	−3.92				
		−28.9	−4.10				
		−27.0	−4.45				
		−25.0	−4.79				

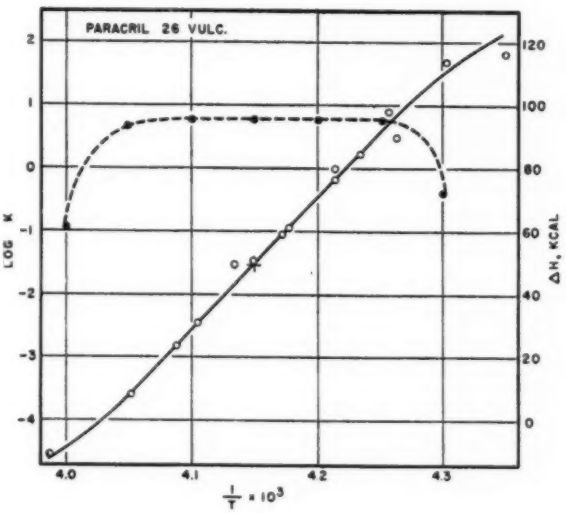


FIG. 9. Temperature dependence of characteristic relaxation time and of apparent activation energy for Paracril 26 gum vulcanizate.

the table are probably the result of incorrect temperature measurement during the earlier runs.

The solid lines in Figs. 5–8 are the Gauss error integral curves which best fit each master curve. These Gaussian master curves should be distinguished from the true master curves defined by the experimental points. The portions of these curves which are shown in Figs. 1–4 were obtained by sliding the entire curve to the left or right by the amount log$K$ taken from Table I. The solid portions of the curves in Figs. 1–4 are portions of the Gaussian master curves, while the broken lines are portions of the true master curves.

In Figs. 9–12 are shown the temperature dependence of log$K$ and of the apparent heat of activation for the four copolymers as measured by the Arrhenius equation

$$H_{\text{Act.}} = 2.303R \frac{d \log K}{d(1/T)}.$$

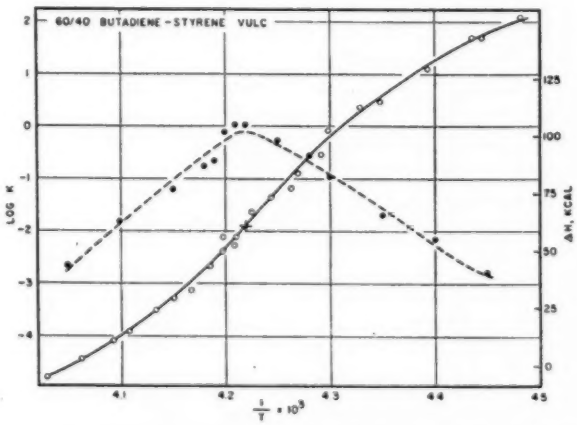


FIG. 10. Temperature dependence of characteristic relaxation time and of apparent activation energy for 60/40 butadiene-styrene copolymer gum vulcanizate.

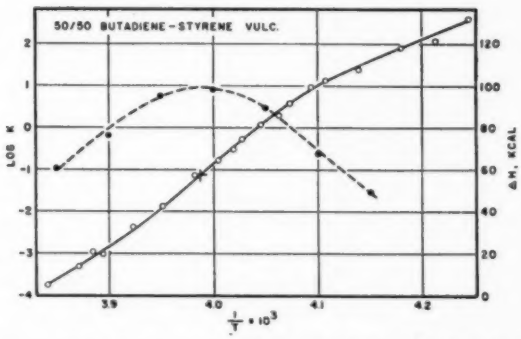


FIG. 11. Temperature dependence of characteristic relaxation time and of apparent activation energy for 50/50 butadiene-styrene copolymer gum vulcanizate.

In all four cases, $H_{\text{Act.}}$ goes through a maximum (rather broad for Paracril 26), so that the definition of T_d proposed in paper I may still be applied. The speculation that K_d was the same for all polymers (about 2 minutes)

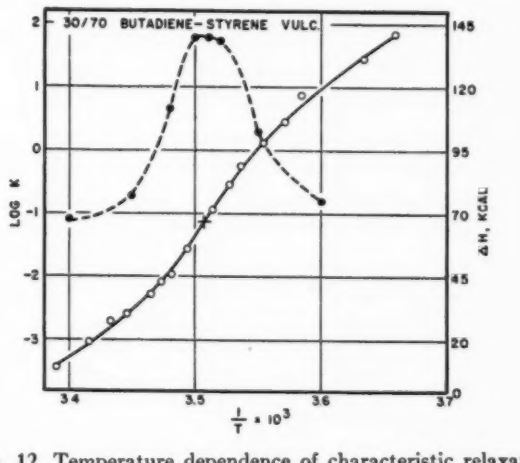


FIG. 12. Temperature dependence of characteristic relaxation time and of apparent activation energy for 30/70 butadiene-styrene copolymer gum vulcanizate.

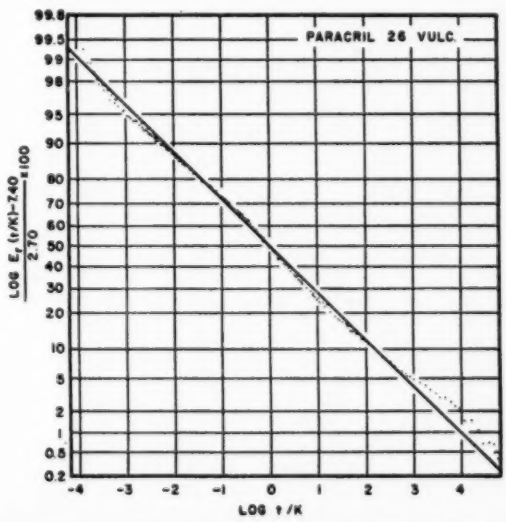


FIG. 13. Master stress-relaxation curve for Paracril 26 gum vulcanizate on normal probability graph paper.

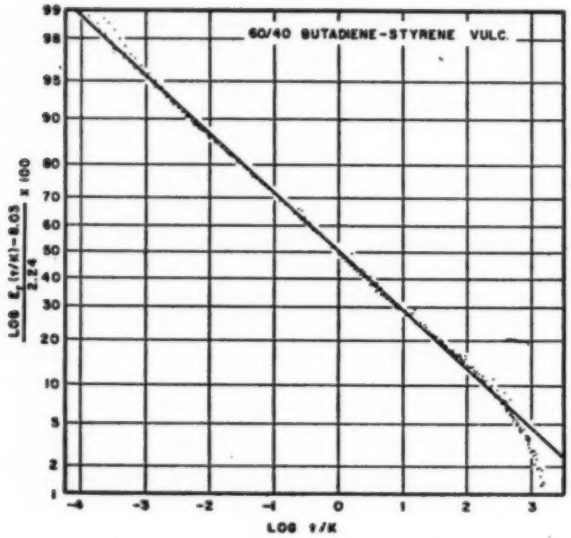


FIG. 14. Master stress-relaxation curve for 60/40 butadiene-styrene copolymer gum vulcanizate on normal probability graph paper.

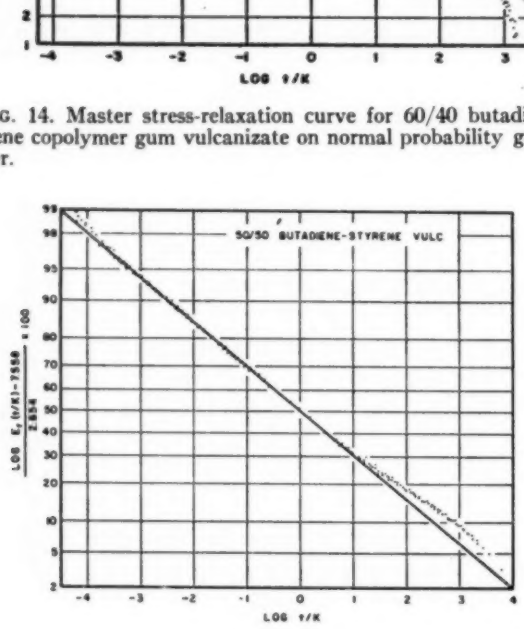


FIG. 15. Master stress-relaxation curve for 50/50 butadiene-styrene copolymer gum vulcanizate on normal probability graph paper.

has been borne out only approximately. The new data include K_d ranging from about 40 seconds to 4.5 minutes. (For dynamic measurements on an unvulcanized GR-S type of copolymer, K_d was found to be as low as 0.001 second.[5])

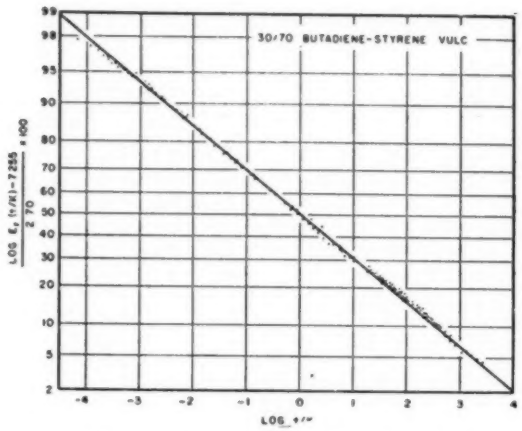


FIG. 16. Master stress-relaxation curve for 30/70 butadiene-styrene copolymer gum vulcanizate on normal probability graph paper.

Using the 30/70 copolymer as an example, the process of evaluating the parameters was as follows. The parameter C of Eq. (1) was readily determined from an enlarged plot of the data in Fig. 8, since it was merely the ordinate of the true master curve at its inflection point. The parameter h was provisionally evaluated from the slope of a plot (not shown) of

$$\log\left[-\frac{d\log E_r(t/K)}{d\log t/K}\right] \textit{ versus } (\log t/K)^2,$$

since the slope is equal to $0.4343\,h^2$. Next the parameter B was evaluated from the y intercept of this plot, which is equal to $\log(2Bh\pi^{-\frac{1}{2}})$. Then $\log E_1(=C+B)$ and $\log E_2(=C-B)$ could be found. At this stage an immediate check on the calculations could be made by observing whether the data of Fig. 8 approached these limits asymptotically. Sometimes minor adjustments in B and C were made at this point.

A better evaluation of h was then obtained by plotting, on normal probability paper,

$$\frac{\log E_r(t/K)-\log E_2}{\log E_1-\log E_2}\times 100 \textit{ versus } \log t/K.$$

These plots (which are straight lines if Eq. (1) holds) are shown for the four copolymers in Figs. 13–16. h was

TABLE II. Parameters of reduced equation for viscoelastic behavior of amorphous polymers.

Polymer	C	B	$\log E_1$	$\log E_2$	h	T_d, °K	hT_d	$\log K_d$
Polymethyl methacrylate	8.85	1.50	10.35	7.35	0.31	384	119	−1.5
Paracril 26	8.75	1.35	10.10	7.40	0.415	241.0	100	−1.54
GR-S	8.84	1.40	10.24	7.44	0.45	220	99	−1.50
60/40	9.15	1.12	10.27	8.03	0.40	237.1	95	−1.94
50/50	8.885	1.327	10.212	7.558	0.364	250.8	91.5	−1.12
30/70	8.605	1.35	9.955	7.255	0.36	285.1	102	−1.13

[5] Zapas, Shufler, and DeWitt (unpublished data).

found from the fact that

$$I_1 - I_{50} = I_{50} - I_{99} = 1.645/h, \quad (4)$$

$$I_{10} - I_{50} = I_{50} - I_{90} = 0.946/h, \quad (5)$$

where $I_1 = x$ intercept at 1 percent ordinate, etc.

The parameters for the six polymers studied so far are summarized in Table II. The near constancy of the product hT_d is noteworthy, although no explanations can be advanced.

REDUCED EQUATION FOR VISCOELASTIC BEHAVIOR OF AMORPHOUS POLYMERS

A reduced form of Eq. (1) may be obtained by plotting

$$[\log E_r(t/K) - C]/B \textit{ versus } h \log t/K.$$

In this form, the Gaussian master curves of all six polymers would become identical. There remains to be considered the reduced-temperature dependence of the reduced characteristic relaxation time. Equation (2) indicates that this should be done by plotting $h \log K_R$ *versus* T_R. Since we have plotted $\log K$ *versus* $1/T$ previously, we chose to plot $h \log K_R$ *versus* $1/T_R$ for all six polymers in Fig. 17. Clearly, the reduced-temperature dependence of the reduced characteristic relaxation time follows Eq. (2) for all the polymers, although the locus of points may more properly be considered a band rather than a single line. This suggests that $f(T_R)$ in Eq. (2) should not be approximated by $-36(T_R - 1)$, but may perhaps be better given in graphical form.

SECOND-ORDER TRANSITION REGION OF BUTADIENE-STYRENE COPOLYMERS

The most obvious trend with composition which appears in Table II is the rise of the distinctive temperature T_d with increasing styrene content of butadiene-styrene copolymers. A theory has been proposed by Gordon and Taylor[6] to account for the composition-dependence of the volumetric second-order transition

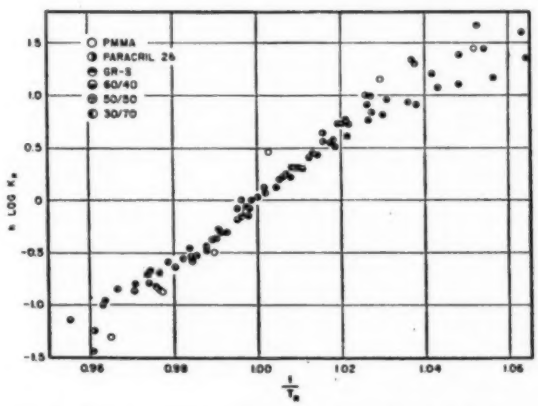


FIG. 17. Reduced-temperature dependence of reduced characteristic relaxation time for six polymers.

[6] M. Gordon and J. S. Taylor, J. Appl. Chem. (London) **2**, 493 (1952).

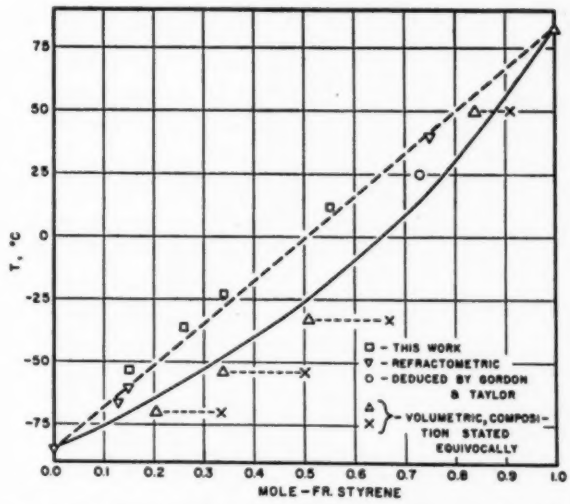


FIG. 18. Second-order transition temperatures of butadiene-styrene copolymers. Solid line is theoretical curve predicted by Gordon and Taylor (reference 6).

temperature T_g of copolymers in terms of the volume occupied by each monomer residue in the glassy and rubbery ("liquid") regions. This theory predicts that the variation of T_g with composition (expressed as mole-percent of monomer units) in butadiene-styrene copolymers should follow the solid curve in Fig. 18.

The four values of T_d found in this work are shown in Fig. 18 as squares. Clearly, they nearly follow a simple straight-line relationship (shown dashed). The remaining points are from the literature, and were cited by Gordon and Taylor[6] in support of their theory. The four pairs of points connected by dashes were determined volumetrically by Ueberreiter,[7] who reported compositions in terms of mole-ratios of components (crosses). Gordon and Taylor evidently plotted the compositions on the assumption that Ueberreiter's figures referred to weight-ratios (upright triangles). The point shown as a circle was calculated by Gordon and Taylor from density data at 25°C in accordance with their theory. The inverted triangles are T_g determined refractometrically[8,9] or by heat-capacity measurements.[9] With the exception of Ueberreiter's points and the calculated point, the straight-line relationship appears to be reasonably valid. A cursory search of the literature reveals no additional data on this series of copolymers, except for confirmation of the approximate value of T_g in the GR-S region.

ACKNOWLEDGMENT

The authors wish to express their sincere appreciation for the support rendered by the Organic Materials Branch of the U. S. Office of Naval Research. This study was sponsored by the U. S. Office of Naval Research, Contract No. N6 onr-27021.

[7] K. Ueberreiter, Z. angew. Chem. **53**, 247 (1940).
[8] R. H. Wiley and G. M. Brauer, J. Polymer Sci. **3**, 455 (1948).
[9] Wiley, Brauer, and Bennett, J. Polymer Sci. **5**, 609 (1950).

The Rheological Unit in Raw Elastomers

M. MOONEY AND W. E. WOLSTENHOLME
United States Rubber Company, General Laboratories, Passaic, New Jersey

(Received April 12, 1954)

Reasons are adduced for suspecting that in a raw elastomer subjected to continuous shear there exist groups of molecules which are held together by entanglement or by thixotropic attachments and move and rotate essentially as elastic solid bodies. Such coherent molecular groups are called rheological units.

If an elastomer in continuous shear consists essentially of closely-packed rheological units, such units, rolling on each other, will transfer a rubber soluble dye across the shear planes as the ink rollers transfer ink in a printing press. This predicted phenomena has been observed with a Mooney viscometer fitted with a smooth rotor, the face of which is covered with a thin film of dried cement containing a dye in high concentration. From the velocity of the color front a rough value can be computed for the mean diameter of the rheological units. The diameters for a series of different elastomers lie in the range from 1 to 35 microns.

INTRODUCTION

IN the generally accepted activation theory of viscous flow as developed by Eyring[1] and his many followers, the activated moving units in a flowing liquid are either the molecules of the liquid, or the chain segments of the molecules if the material is a high polymer. However, there are both theoretical and experimental reasons for believing that, at least in many cases, there are coherent masses containing many molecules of the liquid, which move and rotate essentially as rigid units. In organic liquids of low molecular weight such units are suggested by the x-ray evidence of cybotaxis.[2] Glasses are generally supposed to contain regions of higher and lower order in atomic arrangements[3] the regions of higher order being the disperse phase and presumably exhibiting some rigidity and permanence at the lower flow temperature of the glass. Any colloidal material which exhibits thixotropy must pass through intermediate stages of breakdown in which some fragmented thixotropic structures remain suspended in the more liquid phase.

FIG. 1. Cross-current transfer by rheological units rotating in simple shear. Upward penetration of dye from lower surface is proportional to mean particle diameter, velocity of shear, and total time sheared.

On the basis of the foregoing discussion it would be expected that raw elastomers, which are all highly thixotropic systems, would in their behavior exhibit some evidence of rheological units. Such evidence exists and is most obvious in the behavior of raw elastomers in milling and calendering operations. When a piece of cold raw rubber is placed on a two-roll mill, the rubber at first tears and shreds as it passes through the nip. As the batch warms up, the sheet becomes smoother and more continuous; but even when fully warmed up the sheet still shows more or less roughness, depending on the nature of the rubber. A calendered sheet of gum rubber likewise shows more or less irregularity; and the shrinkage after removal from the calender roll is always irregular and develops waves and thickness variations in the sheet. Such irregular recovery indicates the existence of hard and soft regions of macroscopic dimensions in the rubber.

THE THEORY OF CROSS-CURRENT DIFFUSION

There exists a simple method for measuring the mean diameter of the rheological units in an elastomer, if the units are close packed with a negligible volume of liquid material between them. The method makes use of the fact that all such units will be in rotation in a material subjected to continuous shear.

If a low molecular weight material, a rubber-soluble dye, for example, exists in high concentration along a plane parallel to the planes of shear, the material will be transferred across the shear planes by rotation of the rheological units just as ink is transferred by a series of inking rollers on a printing press. This cross-current diffusion effect is illustrated in Fig. 1. When a dye is being transferred by this process, the velocity v of the color front is

$$v = 2r\omega/\pi. \quad (1)$$

When the units are not uniform, the radius r must be interpreted as a mean radius of the rheological units.

[1] Glasstone, Laidler, and Eyring, *The Theory of Rate Processes* (McGraw-Hill Book Company, Inc., New York, 1941), Chap. X.
[2] G. W. Stewart and R. M. Morrow, Phys. Rev. **30**, 232 (1927).
[3] Zachariasen, J. Am. Chem. Soc. **54**, 3841 (1932). G. Hägg, J. Chem. Phys. **3**, 42 (1935). Warren, Phys. Rev. **45**, 657 (1934).

By the classical theory of the deformation of continuous media, the angular velocity of rotation is expressible in terms of the rate of shear $\dot{\gamma}$ by the equation

$$\omega = \dot{\gamma}/2. \quad (2)$$

EXPERIMENTAL METHOD

The Mooney viscometer is employed for measurements of the size of the rheological units indicated by the above simple theory. The top surface of a smooth rotor is painted with a rubber cement containing a rubber soluble dye,* and allowed to dry overnight at room temperature. The dried film is less than 0.001 in. in thickness, and the concentration of the dye is roughly 300 mg per gram of dry rubber.

Preformed samples are used to minimize the dye penetration due to flow caused by closure of the viscometer. The sample is preformed in the viscometer at 100°C for several minutes with a spare unpainted rotor. The preformed sample is carefully cut off the preforming rotor and then placed in the viscometer with the painted rotor. After one minute preheating at 100°C the sample is sheared at a fixed rate for a specified time. Sample and rotor are then removed from the viscometer and kept in dry ice until the sample is sectioned for photographing. The photography must be carried out promptly after warming the sample, to minimize thermal diffusion of the dye.

Radial sections of the sample approximately 1 mm thick are mounted between two microscope slides with Arochlor 1262 as the mounting fluid. Photographs by direct light transmission are taken at 5× magnification. The portion of the sample containing dye is opaque and is black on the photographs.

Theoretically, the color front in a sample sheared with the disk rotor should be conical, as shown in Fig. 2. The angle α between the cone generator and the rotor surface, is easily shown from previous equations to be

$$\alpha = 2rN/h. \quad (3)$$

In obtaining this equation, $\dot{\gamma}$ was expressed in terms of time, the radial distance, and N, the total number of rotor revolutions, and h, the distance between the stator and the rotor face. The equation involves the approximation $\tan\alpha \doteqdot \alpha$.

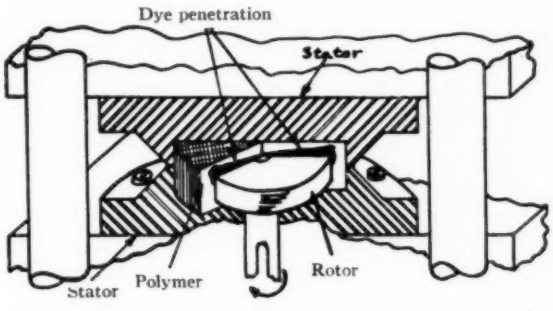


FIG. 2. Schematic of Mooney viscometer chamber showing conical color front along radial section.

* "Hysol Red P or Hysol Green B," Patent Chemicals, 335 McLean Boulevard, Paterson 4, New Jersey.

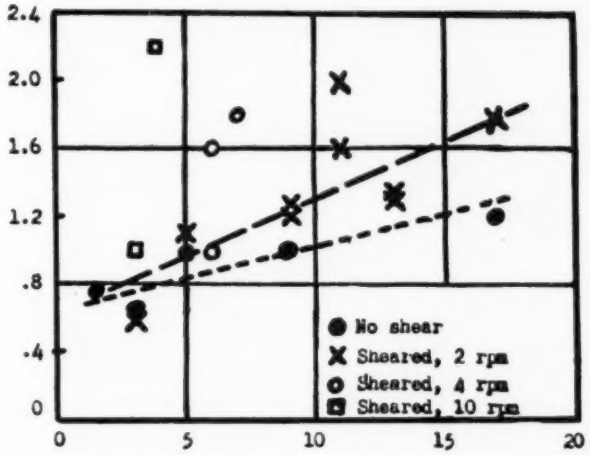


FIG. 3. Total time in viscometer at 212°F, minutes. Effect of shear diffusion of dye at center of sample.

The apex of the cone should lie in the face of the rotor. This ideal condition has never been obtained in experimental work, since the disturbance and flow during closure of the viscometer and the small thermal diffusion during the preheat period contribute to the upward displacement of the dye boundary before the sample is sheared. In Fig. 3 the effect of shear on diffusion of dye at the center of the sample is shown for several shearing conditions. The penetration at no shear increases in a regular manner with total time in the viscometer at 212°F, but the extrapolation to zero time has a positive intercept that indicates the magnitude of the dye penetration due to closure. The measurements on samples sheared at 2 rpm show similar results, with extrapolation to zero time giving practically an identical closure penetration. The increased penetration at the center for sheared samples is attributed to transfer of the dye by rheological units rotating in the vicinity of the center line.

The disturbing effects discussed here seem to cause only a constant additive dye penetration, without appreciable effect on the angle of the color front after shear.

EXPERIMENTAL RESULTS

Estimates of the average diameter of the rotating units are obtained by measuring the color front slope on the photographs and calculating the diameter from Eq. (3). Although most of the exploratory experiments have been made with Hevea, either pale crepe or smoked sheets, other elastomers have been tested. Table I lists a series of common elastomers and the average size of the rheological unit as found by the simple technique described above.

Figure 4 contains unretouched photographs of sheared samples and a sample with no shear. All three samples subjected to shear show the typical angled color front,

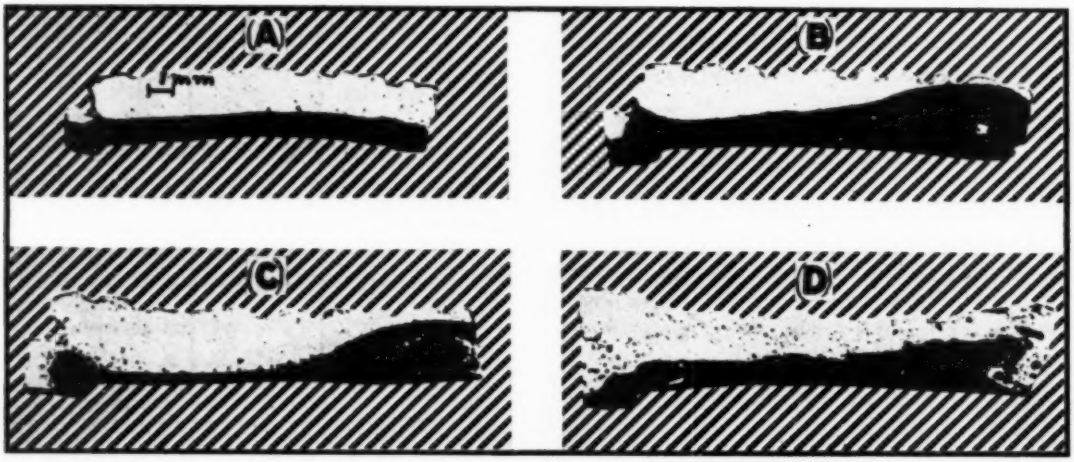


FIG. 4. Rheological cross-current diffusion in several elastomers. Radial sections are mounted with radius increasing to right. Mooney viscometer rotor speed was 2 rpm. (A) Pale crepe (ML4=54) in viscometer 8 minutes but not sheared. Thermal diffusion produces uniform penetration. (B) Pale crepe (ML4=54) sheared 8 minutes. Note increased dye penetration from lower surface as radius increases towards right. Average diameter is 1.43×10^{-3} cm. (C) Butyl (GR-I 15 ML4=48) sheared 10 minutes. Low penetration at left center indicates small rheological units. Large penetration at right is from circulation effects at rotor edge. Average diameter is less than 2.5×10^{-4} cm. (D) Paracril C (ML4=52) sheared 5 minutes. Occasional irregularities in color front slope indicate presence of larger size particles. Average diameter is 3.5×10^{-3} cm.

though the angle in 4(C), butyl (GR-I 15), is very small. The circulation at the edge of the rotor, plainly evident at the right end of Fig. 4(C), is the result of secondary stresses developed in the rubber beyond the edge of the rotor. Such circulation effects must be ignored in measuring the color fronts. The basic theory of secondary stresses in a viscoelastic material has been presented by Mooney,[4] among others, in connection with the Weissenberg effect.

In GR-S polymers White, Ebers, Shriver, and Beck[5] have shown that the presence of gel alters both the processing and physical properties of GR-S compounds. Gel agglomerates in GR-S have been observed in exploratory cross-current diffusion experiments. In Fig. 5 three samples show the turbulent transfer effect attributed to disruption of shear stream lines by large agglomerates of gel. When the sample has no measurable gel the dye transfer by rheological diffusion produces the typical smooth angled color front as seen in 5(D). The gel content in samples of 5(C) and 5(B) were measured by the standard test of determining the benzol insoluble portion remaining on a 50-mesh stainless steel screen.

TABLE I. Average size of rheological units found in common elastomers. All results obtained at a shear rate of 2 rpm.

Polymer	Average diameter of rheological unit	Time sheared	Mooney viscosity ML4
Pale crepe	3.1×10^{-4} cm	20 min	20
Pale crepe	5.0	15	34
Pale crepe	8.3	12.5	35
Pale crepe	14	8	54
Pale crepe	26	5	76
Butyl, GR-I 15	2.5	10	48
Butyl, GR-I 17	2.5	10	69
Hot GR-S	5.7	16	26
Cold GR-S (X-101)	20	6	29
Paracril B	13	10	46
Paracril AJ	23.0	10	44
Paracril C	36	5	52

DISCUSSION

It is to be admitted that the method of computing rheological unit diameters presented in this paper is crude and, furthermore, that there is no obvious way to refine the analysis and improve the computations. On the other hand, the computed diameters are presumably correct as to order of magnitude and relative rating.

While the standard Mooney viscometer has been successfully used in the work presented here, the dimensions of this instrument are not the best for this purpose. A similar apparatus built on a larger scale would be better and, in fact, would be necessary for quantitative measurements on samples like that shown in Figs. 5(A), (B), and (C). A design change that would decrease or eliminate the edge circulation would also be a valuable improvement; but none of the minor changes in cavity shape tried in our own work were successful in accomplishing this purpose.

The method described here has been used successfully with rubber mixes containing white fillers. Mixes containing carbon black have not given satisfactory results with any diffusing material tried. Barton and Ganzhorn[6]

[4] Mooney, J. Coll. Sci. 6, 96 (1951).
[5] White, Ebers, Shriver, and Breck, Ind. Eng. Chem. 37, 770 (1945).
[6] B. C. Barton and G. H. Ganzhorn, "Chemistry of carbon black dispersion," paper presented at Los Angeles Meeting of the American Chemical Society, 1953.

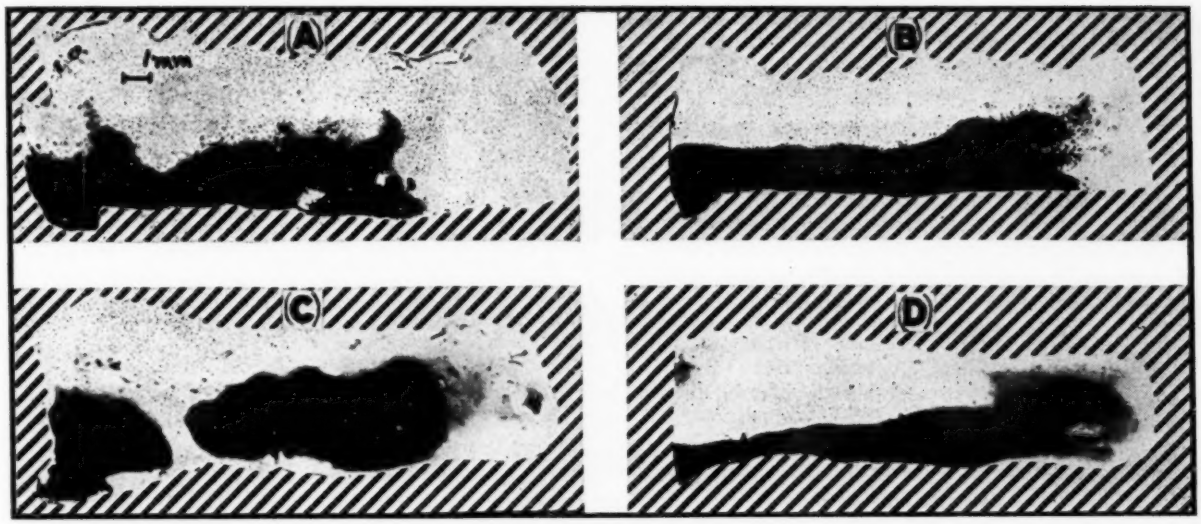


FIG. 5. Effect of gel in rheological cross-current diffusion. Radial sections are mounted with radius increasing to right. Mooney viscometer rotor speed was 2 rpm. (A) Hot GR-S 122°F (ML4=45) sheared 5 minutes. Nonuniform color penetration is attributed to large agglomerates that remain intact and rotate under shear. (B) Hot GR-S broken down to ML4=26. Color front regularity shows agglomerates have been broken up during this type of breakdown. (C) Cold GR-S (X-101) broken down to ML4=30. Large agglomerates are present after this type of breakdown. Measured gel was 15 percent. (D) Cold GR-S (X-101) broken down to ML4=29. Note absence of agglomerates in this special type of breakdown. Measured gel was 0 percent. Rheological unit diameter averages 2×10^{-3} cm.

have suggested that rheological units may influence carbon black mixing.

The successful experimental demonstration of gross rheological units in elastomers suggests a search for such units in other flowing systems. However, the condition that the diffusing molecules must be small in comparison with the rheological units themselves makes the problem difficult. Perhaps heat conduction is a tool that would work with low molecular weight liquids. For example, if there are rheological units in water larger than the H_2O molecules, then the thermal conductance between the two cylinders of a Couette viscometer would be increased when the cylinders are kept in relative rotation.

The authors are pleased to acknowledge their indebtedness to George F. Schrappel for most of the experimental work in the present paper.

Further Dynamic Investigations on Polymers

WLADIMIR PHILIPPOFF
The Franklin Institute Laboratories for Research and Development, Philadelphia, Pennsylvania
(Received April 12, 1954)

Using the vibration tester of The Franklin Institute, a check of the proportionality between stress and strain in dynamic testing disclosed that a proportionality exists when heating is eliminated. A comparison of values of the dynamic test and steady-state experiment in the same range of rates of shear and shearing stress showed that the behavior is basically different; whereas in the dynamic test the viscosity is independent of both amplitude and shearing stress, in the steady-state test experiment a very high dependence exists. The curve of viscosity *vs* rate of shear (D) is very similar to the one of dynamic viscosity *vs* frequency (f). A frequency shift of the order of two is involved in comparing D and f. The application of Ferry's method of reduced variables to a polyisobutylene solution, liquid polyesters, and solid plastics showed its validity. A precision test on the polyester showed that the temperature shift, $T\rho/T_{0\rho_0}$, is necessary to make experimental values at different temperatures fit. The reduced curve of G' *vs* frequency for different plastics has an S shape with different slopes in a very wide range of frequency around 10^{24}.

INTRODUCTION

IN a previous publication,[1] the principle and description of the vibration tester of The Franklin Institute was given. This tester allows us to measure mechanical properties, Shear modulus G' and Lon modulus G'', in a range of frequency of about one to one million in a wide range of stiffnesses. In this paper, we will present some further investigations that test the validity of the assumptions made in applying the theory to dynamic measurements. Furthermore, a rigorous test of the validity of Ferry's method of reduced variables was made on materials whose dynamic viscosity at small frequencies can be determined.

PROGRAM OF EXPERIMENTS

The following points were chosen for experimental verification:

(a) The proportionality between stress and strain in dynamic testing.

(b) The comparison of the values obtained in dynamic testing and those obtained in a steady-state experiment in the rotational viscometer.

(c) The validity of Ferry's[2] method of reduced variables, by measurement at different temperatures.

EXPERIMENTAL PROCEDURE

The vibration tester previously described was modified in that the samples were stressed in a "plunger attachment" that was modeled around the principle of coaxial cylinders with one closed end in which the material is pumped by movement of an inner cylinder as used by W. Philippoff[3] in 1934 and Smith, Ferry, and Schremp.[4] In this instrument, the inner cylinder had a diameter of about 65 mm and length 41.65 mm; the outer cylinder diameter was 77.8 mm. This allows forces of about 10 kg to be exerted on the inner cylinder without cavitation occurring. For this instrument, the relation between the axial forces exerted on the inner cylinder and maximum shearing stresses in the annulus is given by

$$\tau = 876 \cdot (2F) \text{ in dynes/cm}^2,$$

where $2F$ double maximum force in kg. The deformation γ in percent is

$$\gamma = 0.1435 \cdot (2X),$$

where $2X$ double maximum amplitude in microns. We have checked the validity of the formula employed for the calculation of viscosity from the viscous resistance of this arrangement and have found a check with the theory when we used the Bureau of Standards' oil (OB with $212P$ at 25°C). One has to add roughly a thickness of the annulus (5.35 mm) to the length of the inner cylinder to get the effective length used in the calculation.

The stiff materials were measured in a bending arrangement where the ends of a bar 1 in.$\times\frac{1}{4}$ in.$\times 3\frac{1}{4}$ in. were fixed in a steel clamp to insure a horizontal tangent at the ends, and the center of the bar was clamped in a frame that was moved vertically up and down by the usual drive insuring a free length of 20.5 mm on each side between the frame and supports.

With these two attachments to the regular drive, we can enlarge the range of applicability of our vibration tester enormously. The plunger gadget can be used to measure the viscosity of ten poise oil. In the arrange-

[1] W. Philippoff, J. Appl. Phys. **24**, 685 (1953).
[2] Ferry, Fitzgerald, Grandine, and Williams, Ind. Eng. Chem. **44**, 703 (1952).
[3] W. Philippoff, Physik. Z. **35**, 900 (1934).
[4] To the history of the formula for the coaxial cylinders with one closed end and pump action should be noted, that the first experimental formula was given by F. Lawaczek, Z. Deut. Ing. **63**, 677 (1919) and Heinze (thesis Berlin 1925) as cited by G. Barr in his book *Viscometry* (Oxford University Press, London, 1931). The exact formula was published by R. Seeder, Physica **6**, 45 (1940), further by G. W. Nederbragt and J. W. M. Boulhouwer, Physica **13**, 305 (1947), and W. M. Mazee TOM Reel No. 78, p. 2504 as cited by A. Bondi, *The Physical Chemistry of Lubricating Oils* (Reinhold Publishing Corporation, New York, 1951). Smith, Ferry, and Schremp published their formula in J. Appl. Phys. **20**, 144 (1949). The author has developed the expressions for τ and γ, otherwise the calculations are identical with the ones mentioned previously.

ment used, this could easily be extended towards lower viscosities using smaller clearances between the cylinders. With the bending arrangement using thinner specimens the bending modulus of steel has been measured and conforms with the accepted values given in the tables. This means that the range of measurement of the vibration tester begins at $G=10^2$ dynes/cm^2 and goes up to 10^{12}. The properties of the liquids which we will describe were naturally measured with the plunger gadget which can be used up to roughly 10^6 poises whereas the investigations of soft plastics were done with the bending arrangement.

Since the first report was made, the vibration tester was used in a much wider range of temperatures. By using the temperature bath on the vibration tester, we can easily reach a temperature of 90°C. By placing the whole vibration tester in a cold room operating at −23°C and heating the bath, any temperature of the specimen between +25° and −23°C can be maintained. From experience, it was found the temperature equilibrium in plastics requires, under certain conditions, a very long time to attain equilibrium. This does not mean that the low heat conductivity delays the attainment of an equilibrium temperature, but that the properties of the material that are at equilibrium at the particular low temperature require a long time to be attained; times of several days being encountered. This seems to be a general behavior of all materials of very high viscosity at low temperatures.

As this investigation involved the necessity of having the "static" values of viscosity, i.e., the values measured in steady-flow, the liquid samples investigated were measured in a rotational viscometer. This rotational viscometer was modeled around the coni-cylinder viscometer of Mooney and Ewart[5] and allowed us to get absolute values with no end corrections. This viscometer was equally calibrated with the Bureau of Standards' oil, *OB*-10, and similarly as Mooney and Ewart found, the constants of the viscometer were within a few percent of the values calculated from the dimensions. The constants for this viscometer were

$$\text{rate of shear } D=(1.526/T_{10}) \text{ in sec}^{-1},$$

where T_{10} is the time in seconds necessary for the rotor to move by 10°. The shearing stress was calculated as $\tau=26.4\cdot(W-F)$ in dynes/cm^2, where $W=$driving weight in grams and $F=$friction in grams determined from the calibration. F was in the neighborhood of one gram. In a second high shear viscometer, the values are 31.7 and 136.5, respectively. This viscometer could be operated in quite a large range of rates of shear as seen from the experimental data.

MATERIALS USED

The materials used in the following investigations were:

(a) A 20 percent by weight solution of polyisobutylene that the National Bureau of Standards has used in the cooperative program, in decalin. This solution was stored for about six weeks until equilibrium and homogeneity were attained.

(b) Liquid polyesters were used as delivered: paraplex G-25 and an experimental batch of higher molecular weight G-651 of Rohm and Haas.

(c) Plasticized polyvinyl chloride supplied by the Bakelite Corporation for the cooperative program of the ASTM: VU-1920 and VU-1913, whose composition especially concerning the plasticizer used and the molecular weight of the polymer was not known to the author.

(d) Polyethylene "Alathon 1" obtained from E. I. du Pont de Nemours and Company, Inc. was used.

RESULTS

(a) Proportionality between Stress and Strain in the Dynamic Investigations

In all our investigations, it was observed that the hysteresis curve was always a true ellipse. No amount of graphical analysis could detect any deviation from the elliptical shape, which in itself confirms the assumption that stress and strain are proportional to each other. However, a far more rigorous check is possible by investigating the relation between stress and strain changing the amplitude in large ranges. This was performed especially for the polyisobutylene solution using the plunger instrument at two different frequencies. The amplitude was changed by varying the cam setting so as to give measurable values in the whole range of variables. In these particular investigations, a change of not more than 1 to 6 in the amplitude was possible, as higher double amplitudes than 1 mm were not measurable and small amplitudes gave very small forces. For more viscous materials, a check has been done up to an amplitude ratio of 1 to 44. The results are shown in Table I, for the two frequencies. One sees directly from these experiments that there is *no* detectable change in *viscosity* with the change in amplitude or shearing stress, but there is a detectable decrease of about 25 percent in the *shear modulus* with increasing amplitude.

TABLE I. Twenty-percent polyisobutylene in decalin amplitude dependence of properties.

Cam setting 0	Force $2F$ kg	Shear stress τ dynes/cm^2	Amplitude $2X$ μ	Viscosity η poises	Shear γ %	Shear modulus G' dynes/cm^2	Frequency cps
10	3.85	6750	130	5350	18.6	15 610	
20	8.15	14 300	268	5560	38.5	15 830	
30	12.25	21 500	422	5620	60.6	14 650	0.316 cps
40	16.25	28 500	560	5650	80.5	14 620	
50	19.30	33 800	714	5550	102.5	13 200	
60	21.60	37 900	843	5410	121.0	12 300	
10	1.67	2920	138	30 800	19.8	4610	
15	2.50	4390	222	28 700	31.9	4320	
20	3.52	6170	300	29 800	43.0	4520	0.0316 cps
30	5.40	9470	470	30 200	67.5	4160	
40	6.86	12 050	615	29 900	88.5	3860	
60	10.21	17 920	930	29 800	133.5	3700	

[5] M. Mooney and R. M. Ewart, Physics **5**, 350 (1934).

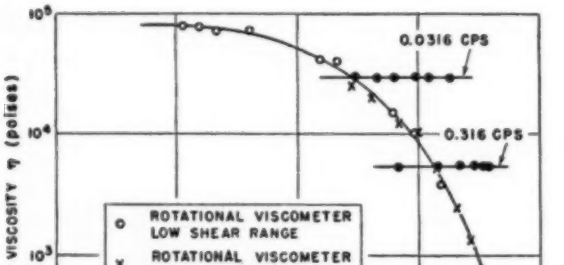


FIG. 1. Dependence of the viscosity on the shearing stress in the dynamic and steady-state experiment.

Obviously $\sin\delta$ increases with amplitude. This behavior is very similar to the one observed with rubbers and other materials. Furthermore, one sees that the viscosity changes by a factor six by changing the frequency from 0.0316 cps to 0.316 cps. At a lower frequency, the value is very near the one of the initial viscosity. This is the maximum viscosity that a non-Newtonian liquid can have under low rates of shear.

(b) Comparison between Dynamic Test and Steady-State Experiments

In Fig. 1, the values of Table I are plotted over shearing stress τ together with values obtained with the rotational viscometer in the same range. We see that these experiments, even if they are made in the same range, are completely incomparable with each other. The steady-state values drop off very fast whereas the dynamic values remain constant. In Fig. 2, the same values are plotted, the steady-state values over rate of shear, and the dynamic values as a function of the frequency f in cps. We see that in this way of plotting, both curves have a much more similar shape up to a decrease of viscosity of about $1 \div 5$. These curves which are a superposition of the reduced curve of polyisobutylene (described in following paragraphs) with the static curve measured at 25°C have been shifted in the

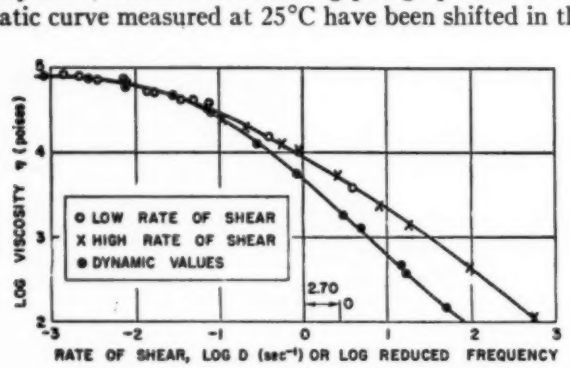

LOW RATE OF SHEAR
HIGH RATE OF SHEAR
DYNAMIC VALUES
2.70
LOG VISCOSITY η (poises)
RATE OF SHEAR, LOG D (sec⁻¹) OR LOG REDUCED FREQUENCY

FIG. 2. Dependence of the viscosity on the rate of shear D, respectively, frequency f, for the dynamic and steady-state experiment.

frequency axis to obtain the best fit of the experiments. The curves had to be shifted by 2.7, whereas the shift to make $\omega = (2\pi f)$ coincides with the rate of shear D should have produced a shift of 6.28. This means that neither f nor ω can be directly compared with D, but a factor of about 2.30 is involved in comparing D and ω. Probably this is not an absolute constant but changes from material to material investigated. The comparison of the properties in this particular experiment was done over very wide ranges of frequency amounting to six decades both in rate of shear and frequency. The values do not check in the whole range but it is very probable that the check is different for other materials than the polyisobutylene used. In general, a similarity of these two curves is unmistakable but one cannot speak of complete coincidence. This is of great importance as the rate of shear cannot be measured up to very high values owing to heat evolution, whereas dynamic measurements, especially using the method of reduced variables, are feasible. They can then be used to extrapolate probable values in the inaccessible range of the "static" experiments. In these experiments, *no* heating of the sample was observed. With other more viscous materials, the experiments were extended into the region when a heating *was* observable. The constancy of η was observed notwithstanding the heating. It is to be noted that dynamic experiments show a faster decrease with frequency than the ones in the steady state. It is important to note that the dynamic values can be both higher and lower than the static experiment performed at the same shearing stress. This seems to be new and completely nonunderstandable.

TABLE II. Polyester *G*-25.

	Viscosity in poises		
	25°C	51°C	73°C
Rotational viscometer	1640	370	144
Vibration tester	1660	373	143

(c) Method of Reduced Variables

Ferry's measurements using the method of reduced variables were essentially confined to gel-like materials such as polyisobutylene and plasticized polyvinylchloride or solution of polymers with an experimentally not attainable dynamic initial viscosity. No values, according to the author, have been published on the materials that behave as non-Newtonian liquids and have a measurable initial viscosity at lower rates of shear or lower frequencies. Some values on low molecular weight polyisobutylene have been reported by T. de Witt and collaborators, on more complicated compounds by H. Kolski (Plasticity Symposium Paper, Providence, R. I., 1953). Ferry's assumption that the modulus has to have a vertical shift determined by $T\rho/T_0\rho_0$ has never been proved experimentally. It was

thought important to have a check on this point with systems which allow independent direct measurements of the initial viscosity; therefore, measurements on the two liquid polyesters were made.

Polyester G-25 measured in the rotational viscometer and in the dynamic tester with the plunger attachment did *not* show any elasticity in the range measured. In this experiment, the values of η determined in the plunger instrument and in the rotational instrument exactly checked. These values are shown in Table II.

Polyester G-651 was investigated at three temperatures; 25°, 51°, and 73°C. A check of the best superposition of the curves at different temperatures showed unmistakably that a small correction due to the temperature-density change *has* to be applied to the experimental values. Measurements with this polyester were extremely well reproducible so that we could, with great certainty, assume this to be generally true insofar as many other substances investigated with less accuracy show the same behavior.[6] The reduced curves for this polyester are shown in Fig. 3.

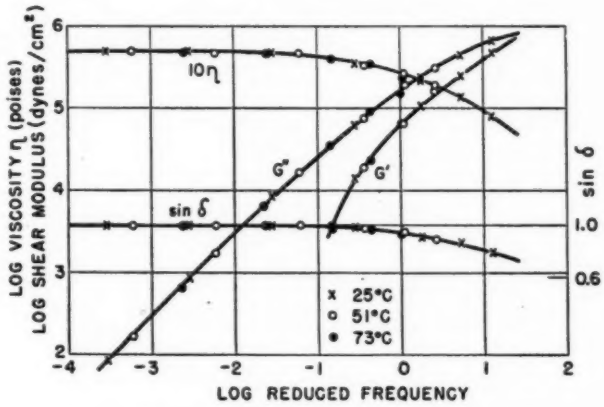


FIG. 3. Reduced variables plot for Polyester *G*-651.

The polyisobutylene solution used in the amplitude test was also investigated at different temperatures. As some difficulties due to the evaporation of the decalin used as the solvent at 50°–73°C were encountered, we investigated this solution at lower temperatures namely; 25°, 0°, and −23°C. A superposition of the values shown in Fig. 4 to a reduced variable curve was possible as seen in Fig. 5. Here again, the vertical temperature shift was used.

Having the bending arrangement available for investigation of solid plastics, these measurements were performed with the solid materials described above. Measurements of Young's modulus calculated from the bending experiments on three solid plastics VU-1913, VU-1920, and polyethylene are shown in Fig. 6. The temperatures used were −23°, 0°, and 25°, 50°, 73°C.

[6] Owing to an experimental accident and the small amount of material available, no exact check with the rotational viscometer was possible. This will be made in the near future with a new batch of material in the whole range of temperatures now available.

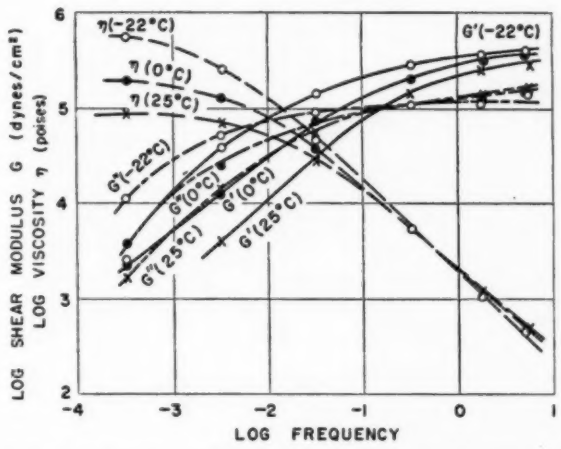


FIG. 4. Experimental values for the dynamic experiments on the 20-percent polyisobutylene solution in decalin.

One sees unmistakably that it should be possible to superimpose these curves by a horizontal shift. This has been made and the curves obtained are shown in Fig. 7 using the temperature shift. They have a typical S shape usual for other soft materials investigated over a large range of reduced frequencies. Especially to be noted is the extreme range of reduced frequencies of about 24 decades for all of the curves investigated. In Fig. 8, the loss angle calculated for the reduced curves is shown. This plot is *linear vs* log frequency. In Fig. 9, the frequency shift a_T *vs* temperature for the three investigated plastics is given. The curves are nearly linear in this plot. These measurements show conclusively as do other measurements (i.e., stress relaxation) that the method of reduced variables *can* be applied with great advantage to solid plastics. Individual curves measured over wide ranges of frequency allow a valid superposition for a total of 24 decades of reduced frequency. It is obvious that even our six isothermal decades of frequency are by far not enough to obtain the S-shaped curve characteristic for the behavior of soft polymers.

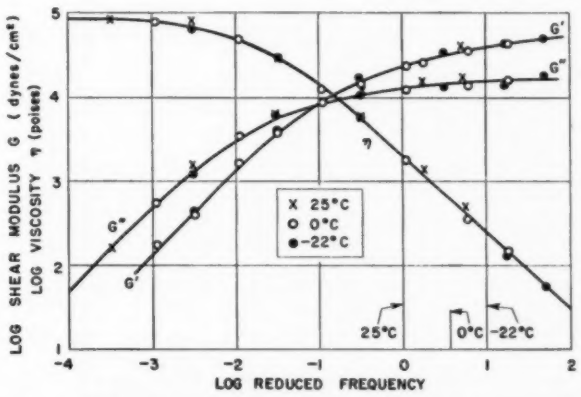


FIG. 5. Reduced variables plot for the polyisobutylene solution. The marked points show the position of $f=1$ at the different temperatures used.

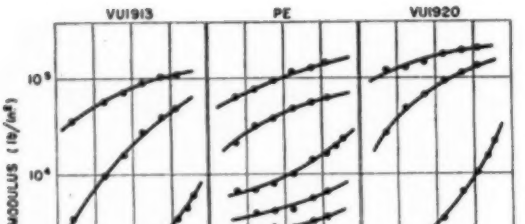


FIG. 6. Dependence of G'' on f for the three solid plastics.

DISCUSSION

(a) Since measurements of dynamic properties of colloidal solutions have been available, one has found *no* evidence of a dependence of the viscosity on the shearing stress. The first measurements by Philippoff in 1934 were performed in the range where the steady-state experiments of Eisenschitz and Rabinowitsch[7] showed pronounced nonlinearity and it was noted that the dynamic behavior was different, in that a proportionality between force and amplitude was present. Ferry's numerous experiments prove the same, even if the amplitudes used were extremely small. The new experiments presented here prove that conclusively. This of course is an important asset of the method of dynamic testing because without an amplitude independence a generalization of the measurements would be practically impossible. There is as yet no pertinent theoretical explanation for this phenomenon. In the accompanying articles by T. de Witt and W. F. G. Swann, attempts have been made in this direction. It is remarkable that whereas the *viscosity* is independent of the amplitude, the *shear modulus does* show a small but definite dependence. As this behavior has been checked on a large number of measurements on other materials, one can assume that this is quite general for non-Newtonian liquids or for that matter for any viscoelastic bodies.

(b) One must accept the evidence that viscosity in dynamic testing is really identical with the viscosity in the steady-state experiment when, because of the particular properties of the materials, Newtonian behavior is possible. For "pure" liquids, this has been known: honey and glycerin were used by Philippoff to calibrate his original instrument. The dynamic values were practically equal to the steady-state ones. The same is the case for the oil *OB*-10. This has been equally proved on very numerous unpublished experiments. The conclusion is that the G'' is really identical with $\eta \cdot \omega$. The complete disagreement of the dependence on shear stress of the dynamic viscosity and the viscosity in the steady-state experiment mentioned in the foregoing has yet to be accounted for. The concurrent shape of the viscosity *vs* rate of shear and viscosity *vs* frequency would tend to show that the reason for this behavior is very similar. It is well known that the velocity or rate of shear in dynamic experiments is proportional to amplitude $X \cdot \omega$. This should compare with the rate of shear D in the steady-state experiment. The splitting up of the product $X \cdot f$ by investigating both the frequency dependence and the amplitude dependence showed that it is f that is the determining quantity not the product $X \cdot f$. It is well known that laminary stationary flow has a rotational component determined by the rate of shear. It is further well known that the rotational component determines a frequency involving the rate of shear and the shape of the particles. This has been used extensively in the calculation of the viscosity of suspensions of ellipsoids. Therefore, it is easy to accept phenomenologically that there is a frequency occurring in laminary stationary flow that should be comparable to the frequency in the dynamic experiment. The ratio of frequency determined in the steady-state experiment to the one in the dynamic test includes the shape of the particles, therefore, the deviation of D and ω as shown in Fig. 1 is understandable. In the dynamic test, we have a phase shift between the force and the deformation: the one measured by the loss angle δ. In the deformation theory of K. Weissen-

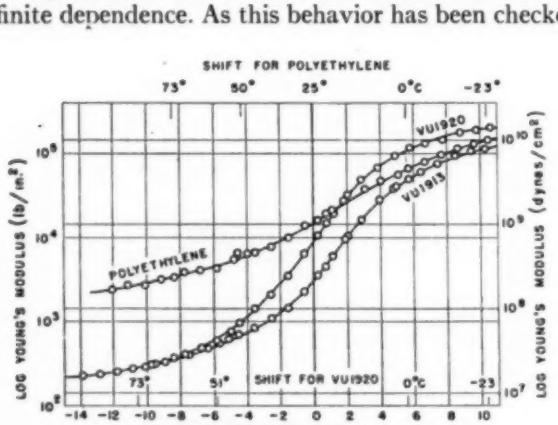


FIG. 7. Reduced variables plot for the three plastics. Shift indicates position of $f=1$ for each temperature.

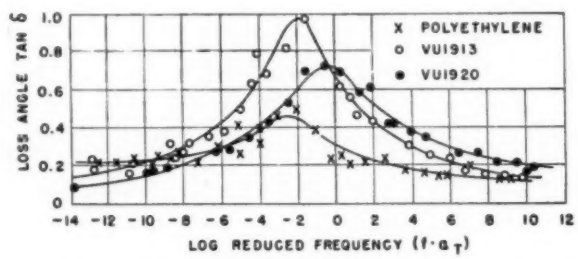


FIG. 8. Loss angle *vs* reduced frequency for the three plastics.

[7] R. Eisenschitz and B. Rabinowitsch, Z. Ber. deut. Chem. Ges. **64**, 2522 (1931).

berg,[8] it has been attempted to correlate the phase shift between the stress tensor and the deformation tensor with the elastic properties of the material. It seems probable that in an as yet unknown way this phase angle is connected with the frequency determined by the rate of shear and δ.

(c) The method of reduced variables receives a confirmation of its validity in a wide selection of materials: solution of polyisobutylene, liquid polyester, and solid materials in a range of reduced variables up to 24 decades. It seemed that there is little doubt as to its wide validity either in the choice of materials, temperature or frequency. This would lead to the confirmation of its basic assumptions: the constancy of the temperature dependence of the viscosity of each partial mechanism in the distribution of relaxation times and the assumption of the kinetic theory of rubber elasticity leading to the temperature shift. Many more experiments of this kind have been performed with exactly the same results. The S-shape curve for the modulus over reduced frequency for the solid plastics investi-

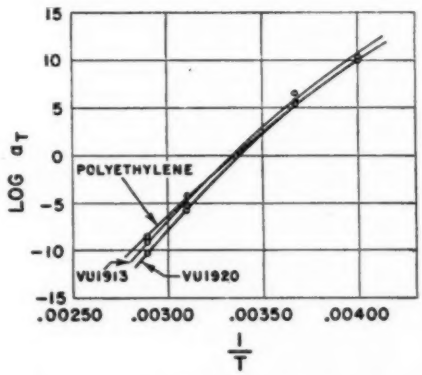


FIG. 9. Frequency shift a_T *vs* reciprocal absolute temperature for the three plastics.

gated has been duplicated with rubber both natural and synthetic, on polystyrene in the transition region[9] and on the plasticized polyvinylchloride by Ferry and Fitzgerald.[10] The presented values especially on the polyethylene show that in order to measure these curves, quite a wide range of frequencies is necessary, much above anything known in other fields of physics. It is not quite clear what his should mean in the sense of a mechanism because a change in frequency of 10^{24} is practically unimaginable. Probably, if we should go to both higher and lower temperatures, this range should be still greater. The occurrence of only one dispersion region for these three plastics is to be noted in confirmation with the other measurements mentioned above. It seems improbable that there could be any measurements done with considerably larger isothermal ranges of frequency. In recent time, there have been some doubts as to the applicability of the method of reduced variables. In isothermal measurements, great care should be taken that there is no stray resonance in the system occurring due to the large change in modulus. On the numerous materials we have investigated, we have found *no* consistent deviation from the method of reduced variables. It is to be noted further that a limiting Young's modulus at lower temperatures and higher frequencies seems to be approached of about 10^{10} dynes/cm^2. It is interesting to note that widely different materials seem to have a common limit under similar conditions of very high frequencies or low temperatures, or for that matter, in the glassy state. One has the impression that most organic materials reach a limit of G' about 10^{10} independent of the chemical composition of the material. Very probably this is caused by the elasticity of the C—C bond. With these materials, no trace of second-order transition could be observed. Measurements of Ferry on the transition of polystyrene were made in a far smaller range of frequencies amounting to four decades for the whole transition as against 20 or more decades necessary here. a_T is essentially the temperature dependence of the initial viscosity. The initial viscosity for these materials is unaccessibly high. The curves are very flat plotted as G' over f. The high temperature coefficient will account for the significant change in G' at different temperatures as especially seen in Fig. 6. It would seem that a high temperature coefficient of a_T and a curve similar to the one measured by Ferry on polystyrene would explain the apparent transition of solid state into the rubbery state when the whole smooth transition region of four decades is determined by only a few degrees of temperature change. There seems to be no measurable discontinuities in the dynamic behavior to be necessary to explain the second order transition at least for the materials investigated. The apparent energy of activation of viscosity flow for the materials investigated are listed in Table III. These values are entirely comparable to the values of about 75–90 kcal/mole for polystyrene at 110–130°C as measured by Ferry.

TABLE III. Apparent activation energies for viscous flow at 25°C

Material	E in $\frac{K \text{ cal}}{\text{mole}}$
Polyethylene	85
VU-1913	104
VU-1920	88

[8] K. Weissenberg, Arch. Sci. phys. et nat. (5) **17**, 1 (1935); see also R. Eisenschitz, Kolloid Z. **64**, 184 (1933).

[9] L. D. Grandine and J. D. Ferry, J. Appl. Phys. **24**, 679 (1953).

[10] E. R. Fitzgerald and J. D. Ferry, J. Colloid Sci. **8**, 1 (1953); E. R. Fitzgerald and J. D. Ferry, J. Colloid Sci. **8**, 224 (1953).

Stress and Strain

Discussion of Dr. Philippoff's Paper*

W. F. G. SWANN†
Bartol Research Foundation of The Franklin Institute, Philadelphia, Pennsylvania
(Received April 12, 1954)

THE simplest relation between stress S, strain σ, and rate of change of strain σ' is given by

$$S=k_0\sigma+\eta_0\sigma', \tag{1}$$

where k_0, the shear modulus of elasticity, and η_0, the viscosity, are absolute constants.

For a sinusoidal relation between σ and t, *viz.*,

$$\sigma=\sigma_0 \sin pt, \tag{2}$$

the strain assumes the form

$$S=S_0 \sin(pt+\psi). \tag{3}$$

Equations (2) and (3) lead to

$$\eta_0=\frac{S_0}{\sigma_0 p}\sin\psi; \quad k_0=\frac{S_0}{\sigma_0}\cos\psi, \tag{4}$$

which relations serve to determine η_0 and k_0 in terms of the measured S_0/σ_0 and the measured ψ.

Equation (1) is linear in σ and σ'.

Now it is possible to conceive of an expression for S which involves not only σ and σ' but all of the higher time derivatives. However, we can, with wide generality, force any such equations into the form

$$S=k\sigma+\eta\sigma', \tag{5}$$

provided that k and η are allowed to be functions of the time derivatives. If k or η are of a form involving ratios of time derivatives and if the order of the time derivatives in the numerators add up to those in the denominators or differ from them by even orders of time derivatives, these expressions for k and σ will not involve the time for sinusoidal strains of the form (2), although they will involve the frequency.

Thus, for example, an expression of the form

$$\left(\frac{\sigma''''}{\sigma''}\right)$$

will, for time variations of the form (2), become simply $(-p^2)$.

Expressions for S involving such forms for k and σ will be highly nonlinear in general, and the relation between S and the time will in general be very complicated functions of the time. For sinusoidal variation of σ, however, they will masquerade as linear relations in the form (5), but with k and η being functions of the frequency.

Now with this understanding, no one can prevent our going through the process of solving (5) for S as a function of the time, and we, of course, get expressions like (4), with k and η replacing k_0 and η_0, but no longer with the limitation of k and σ to being absolute constants, the same for all solutions. We shall have to admit that they can be functions of the frequency, and in determining k and η from equations like (4) through the measurement values of ψ and S_0/σ_0, we shall find that k and η have different values for different frequencies. We shall, in fact, find that η and k are functions of the frequencies and not independent thereof, as they would be if (4) had been derived from an equation like (1), where k_0 and η are assumed to be absolute constants from the start.

Quantities k and η, obtained from experimental measurements based on expressions of the form (4), may in a crude way be regarded as coefficients of shear elasticity and viscosity, now regarded as functions of frequency, and it is customary to so regard them.

However, on the basis of a general equation of the form (5), with k and η regular as functions of higher derivatives, we can ask such questions as to the value of η when the motion is entirely different in form from the sinusoidal. We can make σ' constant with the time, and by measuring the stress when a steady state has been attained, inquire as to the value of η for that case and we can inquire how the value, which we shall call η_∞, depends upon σ'.

With the foregoing considerations in mind, we can start afresh and, supposing that, for η_∞, we have found an expression of the form[1]

$$\eta_\infty=f(\sigma')_\infty, \tag{6}$$

and that for sinusoidal variation we have found

$$k=f_1(p); \quad \eta=f_2(p). \tag{7}$$

We may ask whether it is possible to find an equation of the form (5) which will lead to the experimentally found relation of the form (6) and (7).

* A digest of this paper was presented as a discussion to Dr. Wladimir Philippoff's paper entitled: "Further dynamical investigations on polymers," read before the Society of Rheology at the Hotel New Yorker in New York, New York, October 31, 1953.

† Director, Bartol Research Foundation of the Franklin Institute, and a Senior Advisor for the Franklin Institute Laboratories for Research and Development.

[1] Of course, at $t=\infty$, k_∞ for the steady state must necessarily be zero, since the strain tends to infinity.

Consider an equation of the form[2]

$$S=-\left(\frac{\sigma''}{\sigma''''''}\right)^{\frac{1}{4}} f_1\left[\left(\frac{\sigma''''''}{\sigma''}\right)^{\frac{1}{4}}\right]\sigma''$$
$$-\left(\frac{\sigma''}{\sigma''''''}\right)^{\frac{1}{4}} f_2\left[\left(\frac{\sigma''''''}{\sigma''}\right)^{\frac{1}{4}}\right]\sigma'''$$
$$+\sigma' f_3(\sigma')\left[1-\frac{\sigma''\sigma'''}{\sigma'\sigma''''}\right]. \quad (8)$$

Now while I do not have time to demonstrate the fact, it can be shown with a wide degree of generality that as we approach the limiting value for constant σ', σ''/σ'''' approaches a finite value. It is clear, of course, that for this limiting case σ'' and σ''' are zero, so that the first two terms of this expression disappear if $f_1(\beta)$ and $f_2(\beta)$ are finite when β is finite. It can also be shown that $\sigma''\sigma'''/\sigma'\sigma''''$ approaches zero for the steady state. Thus, we have for this case

$$S=f_3(\sigma')\sigma'$$

and

$$\eta_\infty=f_3(\sigma').$$

However, for the sinusoidal case

$$\sigma''\sigma'''/\sigma'\sigma''''=p^2p^3/pp^4,$$

so that the coefficient of $f_3(\sigma')$ is zero. We are then left with the terms involving f_1 and f_2.

Now for the sinusoidal case $\sigma''/\sigma''''''=1/p^4$,

$$\sigma'''=-p^2\sigma'; \quad \sigma''=-p^2\sigma,$$

we obtain

$$S=f_1(p)\sigma+f_2(p)\sigma'$$

and

$$k=f_1(p); \quad \eta=f_2(p)$$

as desired.

CONSTRUCTION OF A RELATIONSHIP WHICH DEMANDS SIMILARITY BETWEEN f_2 AND f_3

The experiments of Dr. Wladimir Philippoff[3] reveal a rather surprising relation between the functions f_2 and f_3, a relation to the effect that if, for large values of p, we consider corresponding steady-state conditions with velocities σ' given by $\lambda\sigma'=p$, where λ is a constant, then

$$f_3(\sigma')=f_2(\lambda\sigma')=f_2(p). \quad (9)$$

Such a relationship is permitted by (8) but not demanded by it. It is of interest to see how we can specialize our expression for S so as to demand this relation. All that is necessary is to write $f_2(\sigma'\lambda)$ for $f_3(\sigma')$ in (8).

A rather more elegant expression, leading to the same ultimate end point, is

$$S=-\left(\frac{\sigma''}{\sigma''''''}\right)^{\frac{1}{4}} f_1\left[\left(\frac{\sigma''''''}{\sigma''}\right)^{\frac{1}{4}}\right]\sigma''$$
$$+f_4\left[\sigma'\left(1-\frac{\sigma''\sigma'''}{\sigma'\sigma''''}\right)\lambda+\left(\frac{\sigma''''''}{\sigma''}-\alpha_1^4\right)^{\frac{1}{4}}\right]\sigma', \quad (10)$$

with the understanding that $f_4(p^4-\alpha_1^4)^{\frac{1}{4}}$ is defined as $f_2(p)$. In this case, for the sinusoidal case, we obtain

$$S=f_1(p)\sigma+f_4(p^4-\alpha_1^4)^{\frac{1}{4}}\sigma'$$
$$=f_1(p)\sigma+f_2(p)\sigma',$$

while for the steady-state case, we obtain

$$S=f_4(\lambda\sigma')$$

and for large values of $p(=\lambda\sigma')$, $f_4(\lambda\sigma')=f_2(\lambda\sigma')$, as desired.

[2] We could construct an expression involving a smaller number of orders of differentiation than the sixth. However, the form here used tends to avoid certain complications resulting from negative values of the ratio of the derivatives.

[3] W. Philipoff (private communication).

Effect of Rheological Behavior on Thermal Stresses

ALFRED M. FREUDENTHAL
Department of Civil Engineering, Columbia University, New York, New York
(Received April 12, 1954)

Since the conventional elastic analysis of thermal stress problems coupled with limiting creep rates and time-dependent fracture stresses as (inelastic) design criteria, results in design procedures for thermal stresses (in heat exchangers, nuclear reactors, flight structures at supersonic speeds, etc.) of considerable unreality, the effect of various types of rheological behavior (viscoelastic, plastic, work hardening) on the level of thermal stresses is analyzed under simplified assumptions, such as uniaxial stress and polar or cylindrical symmetry. The effect on the thermal stress intensity of the rheological behavior of the material is shown to be very significant, particularly with respect to stress relaxation and the development of residual stresses.

INTRODUCTION

CONVENTIONAL analysis of thermal stresses is based on the assumption of a linear elastic medium with temperature-independent parameters. Since at temperatures at which materials show even a small amount of creep such an assumption is, at best, a crude approximation; the discrepancy between the real thermal stresses and the computed elastic stresses is considerable and increases with time and with temperature. The rheological behavior of real materials generally produces quite significant modifications in the elastic distribution of stresses due to external forces; the average stress intensity, however, is essentially determined by the magnitude of the acting forces. With respect to thermal stresses which are self-equilibrating, it should be noted that not only their distribution but their average intensity itself depends on the rheological behavior of the medium, since they are the result of restrictions of the free temperature expansion and contraction of the medium, imposed by its continuity and by external boundary conditions. The consideration of rheological behavior in the analysis of thermal stresses is, therefore, of still greater urgency than it is in the analysis of load stresses. Equipment for high-temperature service, such as heat exchangers, jet engines, and nuclear reactors, cannot be rationally designed on the basis of elastic thermal stress analysis alone.

In spite of the linearity of the differential equations, solutions of even relatively simple elastic thermal stress problems require, in general, computational efforts of considerable magnitude, particularly for transient temperature distributions. This is due to the fact that the solution of the heat-transfer problem must precede the solution of the thermal stress problem, and that an integral of the former appears as the "load-function" in the differential equations for the latter. The substitution for the elastic stress-strain relations of rheological equations with temperature-dependent material constants produces, in general, systems of nonlinear differential equations with variable coefficients for the stress or strain components. Even for the simplest type of rheological equation and the simplest possible stress conditions, as represented by spherical or cylindrical symmetry, the solution of an inelastic thermal stress problem, while mathematically straightforward, requires elaborate numerical procedures.

1. SPHERE AND CYLINDER WITH SYMMETRICAL TEMPERATURE DISTRIBUTIONS

This is illustrated by the thermal stress equations for the linear viscoelastic (Maxwell) sphere, that are easily derived by combining the equilibrium and the compatibility conditions

$$\partial\sigma_r/\partial r+2(\sigma_r-\sigma_\vartheta)/r=0$$
$$\text{and} \qquad \partial\epsilon_\vartheta/\partial r+(\epsilon_\vartheta-\epsilon_r)/r=0, \tag{1.1}$$

with the linear viscoelastic deviatoric stress-strain relations of the Maxwell body

$$\begin{aligned} 2G(\dot{\epsilon}_r-\dot{e})&=(\dot{\sigma}_r-\dot{s})+(\sigma_r-s)/\tau;\\ 2G(\dot{\epsilon}_\vartheta-\dot{e})&=(\dot{\sigma}_\vartheta-\dot{s})+(\sigma_\vartheta-s)/\tau, \end{aligned} \tag{1.2}$$

where

$$s=(\sigma_r+2\sigma_\vartheta)/3 \quad \text{and} \quad e=(\epsilon_r+2\epsilon_\vartheta)/3, \tag{1.3}$$

and the compressibility relation

$$3K\dot{e}=\dot{s}+3K\alpha\dot{T}. \tag{1.4}$$

σ_r, σ_ϑ, ϵ_r, ϵ_ϑ, G, K, τ, α, and T denote, respectively, the components of radial and circumferential stress, the components of radial and circumferential strain, the shear and bulk moduli, the relaxation time $\tau=\eta/G$ (where η is the coefficient of viscosity), the coefficient of thermal expansion, and the temperature distribution $T=T(r,t)$, obtained as the solution of the heat-transfer equation

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial T}{\partial r}\right)+q/k=qc\dot{T}/k \tag{1.5}$$

for given boundary conditions, rate of heat generation $q(r,t)$, specific density ρ, specific heat c and thermal conductivity k. The resulting differential equation for the radial stress component

$$\left[\partial/\partial t+\frac{1+\nu}{1-\nu}\frac{1}{3\tau}\right]\frac{\partial}{\partial r}\left(r^4\frac{\partial\sigma_r}{\partial r}\right)=\frac{2E\alpha}{1-\nu}r^3\frac{\partial\dot{T}}{\partial r}, \tag{1.6}$$

where $E=2G(1+\nu)=3K(1-2\nu)$ and ν is Poisson's ratio, when integrated twice with respect to r is transformed into the linear differential equation for σ_r:

$$\left[\frac{\partial}{\partial t}+\frac{1+\nu}{1-\nu}\frac{1}{3\tau}\right]\sigma_r = -\frac{2E\alpha}{1-\nu}\left[\frac{1}{r^3}\int r^2\dot{T}dr\right]-C_1/r^3+C_2, \quad (1.7)$$

the integration constants C_1 and C_2 are determined by the boundary and initial conditions. σ_ϑ is then obtained from the first Eq. (1.1)

$$\sigma_\vartheta=\sigma_r+r(\partial\sigma_r/\partial r). \quad (1.7a)$$

Equation (1.7) is valid for constant as well as for temperature- and time-dependent coefficients E and τ. Parallel variability of E and η without change of τ does not change the character of the equation; considering the limited range of the variability of E it has relatively little effect on the stresses. The variability of the relaxation time, on the other hand, changes the character of the differential equation by making its coefficient variable. Since it is of the exponential type

$$\tau=\tau_0\exp(Q/RT^*)/\exp(Q/RT_0^*), \quad (1.8)$$

where Q, R, T^*, T_0^*, and τ_0 denote, respectively, the activation energy, the gas constant, the variable temperature and a reference temperature (in degrees Kelvin), and the relaxation time at the reference temperature, it varies over several orders of magnitude with a relatively small variation of T^*; the effect of such variation on the stresses is considerable.

The form of Eq. (1.7) shows that in a linear viscoelastic (Maxwell) medium temperature stresses are built up only by transient temperature distributions $\dot{T}(r,t)\neq 0$; steady-state distributions $\dot{T}(r,t)=0$ produce no thermal stresses. An initial state of (elastic) thermal stresses existing at $t=0$ will relax exponentially with a relaxation time $3\tau(1-\nu)/(1+\nu)$. For a Maxwell medium with temperature-dependent relaxation time according to Eq. (1.8) the form of the relaxation equation

$$\dot{\sigma}_r+\frac{1+\nu}{1-\nu}\frac{\exp(Q/RT_0^*)}{3\tau_0}\exp(-Q/RT^*)\cdot\sigma_r=0 \quad (1.9)$$

indicates that the stress relaxation at any point is still exponential but that since the relaxation times vary with the radius r because of $T^*=T^*(r)$, the initial stress distribution changes with time.

Relatively simple solutions of the heat transfer problem are obtained for the solid sphere of radius R suddenly cooled from a uniform constant initial temperature T, or heated by linearly increasing the surface temperature, or generating heat at a constant rate. In all these cases the solution has the form

$$T(r,t)=\phi_0(r)+\sum_1^\infty \phi_n(r)\exp(-k\alpha_n^2 t), \quad (1.10)$$

where $\phi_0(r)$ represents the steady-state solution and $\alpha_n=n\pi/R$. Solutions of the form of Eq. (1.10) are obtained only if the effect of the surface- and the boundary-layer heat transfer resistance is neglected, and the temperature of the surface of the sphere is controlled rather than the temperature of the surrounding medium.

From Eq. (1.10),

$$\dot{T}(r,t)=-k\sum_1^\infty \alpha_n^2\phi_n(r)\exp(-k\alpha_n^2 t).$$

Considering that the boundary conditions for a solid sphere are $\sigma_r=0$ at $r=R$ and σ_r finite at $r=0$, the integration constants

$$C_1=0; \quad C_2=\frac{2E\alpha}{1-\nu}\frac{1}{R^3}\int_0^R r^2\dot{T}dr. \quad (1.11)$$

Introducing the abbreviation

$$\Phi_n(r)=\left[\frac{1}{r^3}\int_0^r r^2\phi_n(r)dr-\frac{1}{R^3}\int_0^R r^2\phi_n(r)dr\right], \quad (1.12)$$

the right-hand side of Eq. (1.7) becomes

$$\frac{2E\alpha}{1-\nu}\sum_1^\infty k\alpha_n^2\Phi_n(r)\exp(-k\alpha_n^2 t),$$

leading to the solution for σ_r for the initial and end conditions $\sigma_r=0$ at $t=0$ and $t=\infty$,

$$\sigma_r=\frac{6E\alpha}{1+\nu}\sum_1^\infty \frac{k\alpha_n^2}{[3k(1-\nu)\alpha_n^2/(1+\nu)]-1/\tau}\Phi_n(r) \times\left[\exp(-k\alpha_n^2 t)-\exp\left(-\frac{1+\nu}{1-\nu}\frac{t}{3\tau}\right)\right], \quad (1.13)$$

which is valid for $k\alpha_n^2\geqslant(1+\nu)/3\tau(1-\nu)$.

Equation (1.13) indicates a simultaneous exponential buildup and relaxation of the viscoelastic thermal stresses. The stress distribution is that of the elastic sphere; the stress intensity, however, depends on the difference of exponential functions: the larger the differences between $k\alpha_n^2$ and $(1+\nu)/3\tau(1-\nu)$ the higher the stress intensity. The upper limit of this intensity is that of the elastic body with $\tau\to\infty$; the lower limit is zero and is attained for $k\alpha_n^2=(1+\nu)/3\tau(1-\nu)$. Thus, each term of Eq. (1.13) produces transient stresses which, after reaching a maximum of intensity, decay asymptotically towards zero.

For temperature-dependent relaxation time and $T(r,t)$ according to Eq. (1.10), a solution of Eq. (1.7) can only be obtained by numerical integration.

Yielding of the sphere over a region bounded by the surface $r=\rho$ will occur if the momentary viscoelastic stresses at this surface fulfil the yield condition which, because of the spherical symmetry, can be written in the form

$$(\sigma_r-\sigma_\vartheta)=\pm\sigma_0, \tag{1.14}$$

where σ_0 denotes the uniaxial yield stress. Substituting (1.14) into the first Eq. (1.1) the radial stress in the plastic region of the sphere is obtained:

$$\sigma_{rp}=\pm 2\sigma_0 \ln r+C. \tag{1.15}$$

The integration constant C is determined by the condition that at the free surface of this region $r=a$, $\sigma_{rp}=0$. Hence

$$\sigma_{rp}=\pm 2\sigma_0 \ln(r/a),$$

and

$$\sigma_{\vartheta p}=\pm\sigma_0[2\ln(r/a)-1]. \tag{1.15a}$$

Since the radial stress must be continuous across the viscoelastic plastic boundary $r=\rho$, the momentary location of this boundary can be determined from the condition that $\sigma_\rho=\sigma_{rp}$, where $\sigma_\rho=(\sigma_r)_{r=\rho}$ from Eq. (1.7), or

$$\sigma_\rho=\pm 2\sigma_0 \ln(\rho/a). \tag{1.16}$$

While the stress components in the plastic part of the sphere must fulfill the time-independent yield condition (1.14), they are themselves time dependent because of the time dependence of ρ.

Reversal of the temperature change $\dot{T}(r,t)$ after part of the sphere has become plastic will produce residual stresses that can be obtained by superimposing upon the transient viscoelastic-plastic stresses due to $\dot{T}(r,t)$ a transient viscoelastic stress field caused by $-\dot{T}(r,t)$; the residual stresses will relax according to Eq. (1.9).

For the cylinder with axially symmetrical temperature distributions Eqs. (1.1) are replaced by

$$\partial\sigma_r/\partial_r+(\sigma_r-\sigma_\vartheta)/r=0$$

and

$$\partial\epsilon_\vartheta/\partial r+(\epsilon_\vartheta-\epsilon_r)/r=0, \tag{1.17}$$

while the stress-strain relations Eqs. (1.2) and (1.4) remain valid with

$$s=(\sigma_r+\sigma_\vartheta+\sigma_z)/3 \quad \text{and} \quad e=(\epsilon_r+\epsilon_\vartheta+\epsilon_z)/r; \tag{1.18}$$

the subscript z refers to components in the direction of the cylinder axis.

The stresses can be obtained by direct integration only for the thin circular disk (plane stress with $\sigma_z=0$). By introducing the auxiliary variables

$$(\sigma_r+\sigma_\vartheta)=3s \quad \text{and} \quad \sigma_r-\sigma_\vartheta=f, \tag{1.19}$$

and by substituting them into the Eqs. (1.2), (1.4), and (1.17) the following equations are obtained:

$$[(1+G/3K)\partial/\partial t+1/\tau](\partial f/\partial r+2f/r) = 3G\alpha\partial\dot{T}/\partial r, \tag{1.20}$$

and

$$\partial f/\partial r+2f/r=-3\partial s/\partial r. \tag{1.21}$$

Hence

$$3[\partial/\partial t+3/2(1+\nu)\tau]\partial s/\partial r=-E\alpha(\partial\dot{T}/\partial r), \tag{1.22}$$

and

$$f=-r\cdot\partial s/\partial r+F(t)/r^2. \tag{1.23}$$

$F(t)$ is an integration function to be determined from the initial and boundary conditions.

For the long cylinder (plane strain with $\dot{\epsilon}_z=0$) the combination of the equilibrium equations and the compatibility conditions with the linear viscoelastic stress-strain relations, the compressibility equation, and the condition $\dot{\epsilon}_z=0$ produces two simultaneous partial differential equations for σ_r and σ_z that can be separated only under the assumption of incompressibility; in this case, the equation for σ_r for the Maxwell cylinder can be directly written on the basis of the elastic viscoelastic analogy[1]

$$(\partial/\partial t+1/\tau)\frac{\partial}{\partial r}\left(r^3\frac{\partial\sigma_r}{\partial r}\right)=-6G\alpha r^2\frac{\partial\dot{T}}{\partial r}, \tag{1.24}$$

while the component σ_z is obtained in a similar way from

$$(\partial/\partial t+1/\tau)\left(\sigma_z-\sigma_r-\tfrac{1}{2}r\frac{\partial\sigma_r}{\partial r}\right)=-3G\alpha\dot{T} \tag{1.25}$$

after Eq. (1.24) has been solved.

2. STRUCTURAL MODEL FOR THERMAL STRESS ANALYSIS

The foregoing illustration of the solution of a simple thermal stress problem for the linear viscoelastic plastic sphere suggests that it may be rather difficult to obtain such solutions for any less simple rheological material. On the other hand, the considerable difference in the thermal stress intensity according to the conventional elastic and the linear viscoelastic analysis shows the necessity of a more realistic evaluation of the thermal stresses; the existence in practically all materials of a certain amount of creep at elevated temperatures is quite generally recognized even in the conventional design procedures for elastic thermal stresses by the specification of a limiting creep strain or creep rate as a principal design criterion. Some approximate analysis of the effect of the different types of rheological behavior, at least on the "intensity" of the thermal stresses if not on their distribution, appears, therefore, to be urgently needed. This "intensity" in a 3-dimensional state of stress is more or less arbitrarily defined as that part of the expressions for the stress components which does not depend on the coordinates; for the viscoelastic sphere, for instance, it would be obtained by setting in Eq. (1.13) $\Phi_n(r)=1$.

[1] T. Alfrey, Quart. Appl. Math. 2, 113 (1944).

In a previous paper[2] a simple structural model has been proposed for the analysis of the "intensity" of the thermal stresses in viscoelastic materials of constant and of temperature dependent parameters. This analysis is extended to general rheological behavior and to a systematic representation of the effects of temperature cycles of different duration and amplitude.

The model which has been used in the discussion of problem types of plastic deformation by Hencky and by Prager[3] is presented in Fig. 1. It consists of three bars of cross section $A_1=A_3$, $A_2=A$ and of length $l_1=l_3=l/\sin\phi$, $l_2=l$; it is assumed for further simplification that bars 1 and 3 remain at constant temperature, while bar 2 is subject to a uniformly distributed change of temperature T over its length, and that bars 1 and 3 remain elastic while only bar 2 shows rheological behavior. Bars 1 and 3 thus resist the temperature expansion or contraction of bar 2.

The general equation for the force S in bar 2 at any time t

$$\dot{\epsilon}+\dot{\epsilon}_1 A/A_1\sin^2\phi=\alpha\dot{T}-f(\dot{P}), \tag{2.1}$$

where $\dot{\epsilon}$ and $\dot{\epsilon}_1$ are the strain rates in bar 2 and in the bars 1 and 3, respectively, expressed as functions of the stress $\sigma=S/A$ in bar 2, and P is an external vertical force. The functions $\dot{\epsilon}(\sigma)$ and $\dot{\epsilon}_1(\sigma)$ are one-dimensional rheological equations.

In the case of unlimited linear elasticity of bars 1 and 3 the strain rate $\dot{\epsilon}_1=\dot{S}/AE$, so that Eq. (2.1) becomes

$$\epsilon+\gamma\dot{S}=\alpha\dot{T}-\gamma\dot{P}, \tag{2.2}$$

where $\gamma=1/E_1A_1\sin^2\phi$. It is obvious that for the fully elastic system with $\dot{\epsilon}=\dot{\epsilon}_1=\dot{S}/AE$ the temperature force due to a temperature change in bar 2 is proportional to the momentary temperature difference

$$\dot{S}=EA\alpha\dot{T}/(1+EA\gamma)=c\dot{T}, \tag{2.3}$$

where $c=\alpha AE/\beta$ and $\beta=(1+EA\gamma)$. Beta represents a "constraint-factor" that can vary between 0 (no constraint) and 1 (full constraint); in the three-dimensional case a similar factor depends on the tri-axiality of the state of stress and Poisson's ratio, and varies between 0 and $1/(1-2\nu)$. For joint action of P and T

$$\dot{S}=c\dot{T}-\dot{P}(1-1/\beta). \tag{2.3a}$$

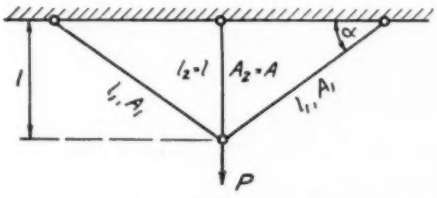


FIG. 1. Structural model for thermal stress analysis.

[2] A. M. Freudenthal, v. Mises Memorial Volume, Academic Press, Inc., New York, 1954.
[3] H. Hencky, Z. angew. Math. u. Mech. 4, 323 (1924); W. Prager, J. Aeronaut. Sci. 15, 337 (1948).

The effect of the rheological behavior of bar 2 on the intensity of the thermal force can now be discussed by introducing in Eq. (2.2) various types of such behavior in terms of the respective functions $\dot{\epsilon}(\sigma)$.

(a) Linear Viscous and Viscoelastic Behavior with Constant Parameters

With $\dot{\epsilon}=S/A\lambda$, where λ is the coefficient of viscous traction, Eq. (2.2) for a viscous bar 2 takes the form

$$\dot{S}+S/\tau_v=c_v\cdot\dot{T}-\dot{P}, \tag{2.4}$$

where the relaxation time of the system $\tau_v=A\lambda\gamma$ and $c_v=\alpha\gamma$. For a viscoelastic bar with creep (Maxwell bar) with $\dot{\epsilon}=\dot{S}/AE+S/A\lambda$ the equation for S has the same form as Eq. (2.4), with different coefficients:

$$\dot{S}+S/\tau=c\dot{T}-\dot{P}(1-1/\beta), \tag{2.5}$$

where the relaxation time $\tau=\beta\lambda/E$.

The solution of Eqs. (2.4) and (2.5) for $\dot{P}=0$ and a temperature change of the type $T=T_0[1-\exp(-t/t_0)]$:

$$S=cT_0\frac{1}{t_0/\tau-1}[\exp(-t/t_0)-\exp(-t/\tau)]$$

$$=cT_0\frac{1}{t_0/\tau-1}[(1-T/T_0)-(1-T/T_0)t_0/\tau] \tag{2.6}$$

is of a form similar to the individual terms of Eq. (1.13) for the stress σ_r in the viscoelastic sphere. The force S increases from zero at $t=0$ towards a maximum S_m that is attained at $t_c=t_0\ln(\tau/t_0)/(1-t_0/\tau)$, and from which it decays exponentially. The value of

$$S_m=cT_0(\tau/t_0)\exp\left[\frac{\ln(\tau/t_0)}{1-(t_0/\tau)}\right] \tag{2.7}$$

decreases with increasing ratio (t_0/τ). For a linear temperature change $T=T_0\,t/t_0$ for $t\leqslant t_0$, the temperature force

$$S=cT_0[1-\exp(-t/\tau)]/(t_0/\tau)$$

$$=cT_0\left[1-\exp\left(-\frac{Tt_0}{T_0\tau}\right)\right]\Big/(t_0/\tau) \tag{2.8}$$

increases at a decreasing rate. When at any time $t_1\leqslant t_0$ the temperature is sustained at its level $T(t_1)$, the force $S(t_1)$ relaxes exponentially with relaxation time τ.

When the temperature change is reversed at $t=t_0$ before the builtup temperature force is relaxed, the decay of this force proceeds simultaneously with the building up of the temperature force according to Eq. (2.8), with a negative value of $c\dot{T}=-cT_0/t_0$. Hence, the temperature force at any moment $t_1=t-t_0$,

$$S(t_1)=S(0)\exp(-t_1/\tau)$$
$$-cT_0[1-\exp(-t_1/\tau)]/(t_0/\tau), \tag{2.9}$$

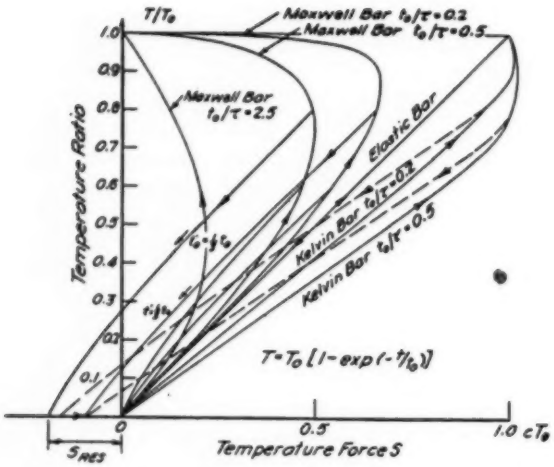


FIG. 2. Relations between temperature and force in bar 2 in terms of the elastic force cT_0, for $T=T_0[1-\exp(-t/t_0)]$.

where $S(0)=S(t_0)=cT_0[1-\exp(-t_0/\tau)]/(t_0/\tau)$. The temperature-force cycle is, therefore, nonsymmetrical; the rates of force reversal are faster than they would be according to Eq. (2.8). Transient residual forces will, therefore, exist for some time after the temperature cycle has been completed, and be the higher the more rapid the decay of $S(t_0)$ and the decrease of temperature. Thus, temperature cycles that are short in relation to τ will be accompanied by fairly parallel cycles of *pulsating* forces; temperature cycles of long duration, however, will produce cycles of force *reversal* that may lead to fatigue fractures. The shorter the duration of the temperature decrease in relation to the temperature application the higher the level of residual forces.

Force-temperature relations $S(T)$ plotted on Figs. 2 and 3 clearly show the differences in both magnitude and sign of the temperature forces in a viscoelastic and in an elastic model.

The deformation of bar 2 is the difference between its unrestricted change of length due to T and its elastic deformation due to S. For the elastic bar, therefore, $\epsilon=\alpha T(1-1/\beta)$; since for the viscous and the viscoelastic bar S is a function of t, the strain $\epsilon(t)=\alpha T[1-S(T)/cT]$, where $S(T)/cT$ is the ratio between the inelastic and the elastic temperature force; assuming, in first approximation, that this term is small in relation to one, particularly after decay of S has taken place, the temperature scale in Fig. 1 can be roughly considered as the scale of inelastic strain $\epsilon_p \doteq \alpha T$. Under this assumption the areas of the $S-T$ loops are roughly proportional to the strain energy dissipated in each temperature cycle.

When a temperature increase to a steady operating temperature T_0 is followed by a slow temperature fluctuation $\pm\Delta T_0$ between the control levels, it is obvious that the thermal force due to T_0 will decay exponentially. The remaining "steady-state" temperature forces will, therefore, only represent the force reversal about zero level due to $\pm\Delta T_0$.

For a viscoelastic bar with after effect (Kelvin bar) $S=AE\epsilon+A\lambda\dot{\epsilon}$ or $\dot{S}=AE\dot{\epsilon}+A\lambda\ddot{\epsilon}$. Combining these equations with the model Eq. (2.2), the equation for the thermal force S is obtained

$$\ddot{S}+\dot{S}/\rho=c\left[\dot{T}/\rho+\frac{\beta}{\beta-1}\ddot{T}\right], \qquad (2.10)$$

where the retardation time $\rho=\lambda(\beta-1)/\beta E$. Assuming $T=T_0[1-\exp(-t/t_0)]$, the thermal force

$$S=cT_0\left[1-\frac{\beta}{\beta-1}\exp(-t/t_0)+\frac{1}{\beta-1}\exp(-t/\rho)\right]. \qquad (2.11)$$

For $t_0=\rho$ and $t=t_0$: $S=cT_0(1-1/e)$; for $t\to\infty$, $S\to cT_0$. For linear temperature change $T=T_0t/t_0$ the force

$$S=cT_0[1+t/t_0-\exp(-t/\rho)]>cT_0 \qquad (2.12)$$

is always higher than the force in an elastic bar. The maximum force at $t=t_0$ for $t_0<\rho$ tends towards $S=cT_0(1+t_0/\rho)$; for $t_0>\rho$ it tends towards $2cT_0$.

Force temperature diagrams for the model with a Kelvin bar have been plotted in Figs. 2 and 3. Comparison with the $S(T)$ diagrams for the Maxwell bar leads to the conclusion that in materials with creep the intensity of the direct temperature stresses is significantly reduced in comparison to the elastic stresses, and residual stresses of opposite sign are built up upon temperature reversal, while in materials with after effect the direct temperature stresses may significantly exceed the elastic stresses, but the residual stresses are relatively unimportant.

(b) Linear Viscoelastic Behavior with Variable Parameters

Because of the very pronounced temperature dependence of the coefficient of viscosity, the equations with constant coefficient (2.2) and (2.3), and their solutions

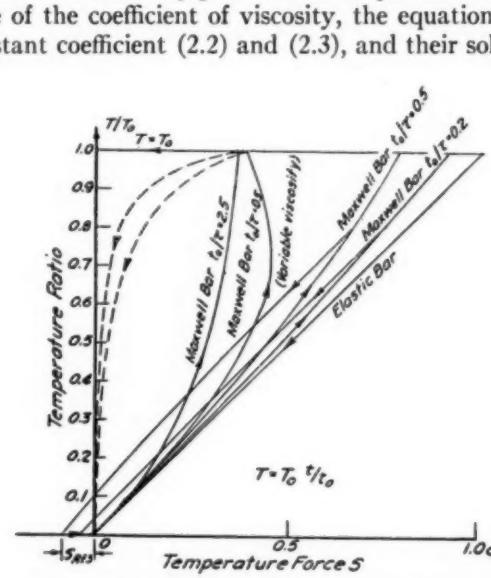


FIG. 3. Relations between temperature and force in bar 2 in terms of the elastic force cT_0 for linear temperature change $T=T_0t/t_0$.

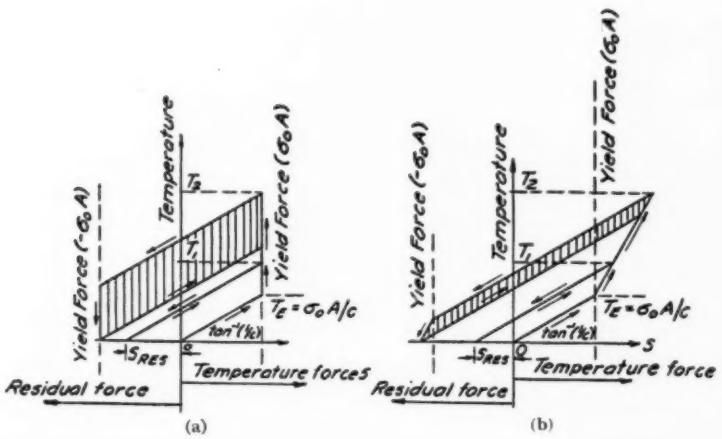


FIG. 4. Relations between temperature and force in bar 2 for linear temperature change and (a) ideal plastic, (b) linear work-hardening behavior of the bar.

represent only a first approximation to the problem of viscoelastic thermal stresses. The introduction of a variable viscosity affects the stress intensity very considerably. Assuming that this variation with temperature can be expressed by Eq. (1.8) and that in first approximation, $(T^*-T_0^*)/T^*=T/T^*\doteq T/T_0^*$, Eq. (2.5) is transformed into

$$\dot{S}+\exp(aT/T_0^*)S/\tau_0=c\dot{T}-\dot{P}(1-1/\beta), \quad (2.13)$$

where $a=Q/RT_0^*$. The solution of this equation for a linear temperature increase $T=T_0(t/t_0)$ for $t\leqq t_0$ and for $\dot{P}=0$ in the form of an infinite series[2] shows that the thermal forces that can be built up in a viscoelastic system are substantially reduced by the temperature dependence of the coefficient of viscosity in the heating part of the cycle. During the subsequent temperature decrease towards the initial temperature the forces decay very rapidly (see Fig. 3).

For certain materials, such as ceramics and glasses, the consideration of λ or τ as a function of temperature alone is not sufficient, since the change of λ with temperature does not occur instantaneously but with a delay; this is probably due to the fact that time is required for the structural changes associated with change of the viscosity to take place in solids and quasi-solids, particularly for materials with low thermal conductivity, and at temperatures considerably below the melting point. A delay function of the form $d\lambda/dt=c_1(\lambda_\infty-\lambda)/\lambda$ has been proposed for glass,[4] where λ_∞ denotes the steady-state value of λ associated with $(T_0^*+T_0)$, λ the value at intermediate temperatures (T_0^*+T) and λ_0 the initial value at T_0^*.

For short times t the relation is approximately $\lambda=\lambda_0(1-c_1t/\lambda_0)$, while for long times

$$\lambda=\lambda_0[1+(\lambda_\infty-\lambda_0)(1-e^{-c_1t})/\lambda_0].$$

Introducing these limiting expressions into $\tau_0=\lambda_0/E$ of Eq. (2.13), it is immediately evident that forces due to temperature changes of short duration are governed by Eq. (2.5) with $\tau=\tau_0$, while forces due to temperature changes of long duration are governed by Eq. (2.13). Hence, whenever the delay in the change of viscosity with temperature is significant in relation to the duration of the temperature change (for instance in quenching operations), heating followed by rapid cooling will produce smaller thermal forces than heating followed by slow cooling.

Because of the linearity of the viscoelastic equations the effect of an applied force can be added to the temperature forces. According to Eqs. (2.4), (2.5), and (2.13), a force due to $P=$const and $\dot{P}=0$ relaxes exponentially. Such relaxation is accelerated with increasing temperature and decreasing relaxation time; the external force is therefore gradually transferred to the slowly relaxing or nonrelaxing parts of the structure, which are the parts that remain at lower temperatures.

(c) Plastic Behavior

The temperature forces in a model with an elastic-plastic bar of yield stress σ_0, the behavior of which is defined by $\epsilon_E=S/AE$ for $S/A\leqq\sigma_0$ and by $\epsilon_P>\epsilon_E$ for $S/A=\sigma_0$, depend on time only as far as the temperature change is time dependent. For a linear temperature change the force in the bar increases linearly until $cT=\sigma_0A$; at the temperature $T_E=\sigma_0A/c$ the bar starts to yield and for $T>T_E$ no further increase of S takes place. On reversal of the temperature at T_0 to zero reference level the force S decreases linearly according to $(-cT)$; the residual force $(\sigma_0A-cT_0)=S_{RES}\geqq(-\sigma_0A)$ makes any subsequent temperature cycle between zero and $T<T_0$ fully elastic. If $(\sigma_0A-cT_0)\leqslant(-\sigma_0A)$ the bar also yields on temperature reversal; every subsequent temperature cycle between zero and T_0 produces "alternating plasticity" which finally leads to failure (Fig. 4a). The number of cycles that can be sustained depends on the strain energy dissipated per cycle.[5] Since for $T>T_E$ the plastic strain $\epsilon_P=\alpha(T-T_E)$, the loop of the $S-T$ diagram is proportional to the dissipated strain energy.

[4] H. R. Lillie, J. Am. Ceram. Soc. **16**, 619 (1936).

[5] L. F. Coffin, "Study of the effects of cyclic thermal stresses on ductile metal," Am. Soc. Mech. Engrs., Paper 53-A-76 (1953).

FIG. 5. Temperature-force cycles in bar 2 for linear work-hardening bar with recovery in the heating part of the cycle.

For a model with a work-hardening bar the behavior of which is defined by $\epsilon_E = S/AE$ for $S/A \leqq \sigma_0$ and $\epsilon_E + \epsilon_P = S/AE + (S-\sigma_0 A)/AH$ for $S/A \geqq \sigma_0$, where $H < E$ denotes the work-hardening modulus, the thermal forces produced by temperature cycles are shown diagrammatically in Fig. 4b. The behavior is quite similar to that of the elastic-plastic bar illustrated in Fig. 4a.

A somewhat different behavior is produced, however, by the assumption that work-hardening effects in the heating cycle either do not exist or are very unstable, while they are fully operative during temperature reversal near the zero reference temperature. Thus, the yield-stress in the heating cycle is either not raised at all during plastic deformation, or the hardening effect is rapidly removed by recovery before the start of the next cycle. Figure 5 illustrates this behavior and shows that when work hardening is stable only in the cooling part of the cycle, the system under repeated temperature cycles tends towards an elastic state with very high residual stresses that will either produce brittle fracture after a few load cycles, or fatigue fracture in the long run.

In Figs. 4 and 5 the diagrams of temperature forces in both the elastic-plastic and the work-hardening bar have been drawn under the assumption that the yield stresses

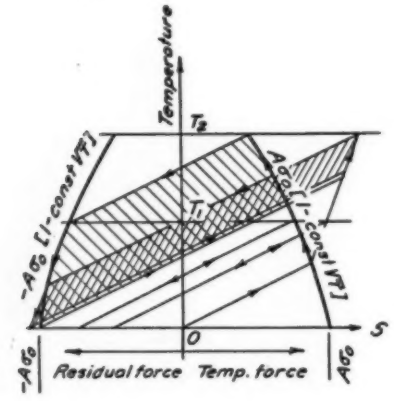


FIG. 6. Effect of temperature dependence of yield stress on temperature-force relations in ideal plastic and linear work-hardening bar 2.

in tension and in compression are the same, and that, moreover, they do not depend on temperature. Figure 6 shows the thermal forces due to linear temperature cycles in a model with an elastic-plastic bar or a work-hardening bar with temperature-dependent yield stress $\sigma_0(1-\text{const}T^2)$.[6] The effect of the decrease with temperature of the yield stress is to increase the residual stresses on temperature reversal.

(d) Viscoplastic Behavior

The behavior of real materials, particularly metals, in the range of high temperatures is a mixture of viscoelastic and viscoplastic behavior, with temperature-dependent coefficient of viscosity and yield limit. The temperature forces in a model with a bar of such mixed behavior can be directly discussed on the basis of the diagrammatic representation used in the viscoelastic and elastic-plastic analysis. Figure 7 illustrates the thermal forces produced by linear temperature cycles in

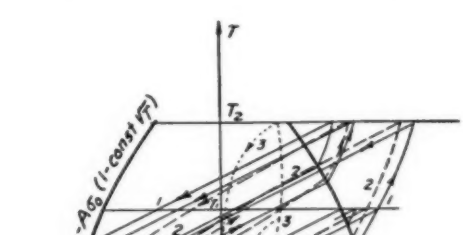


FIG. 7. Temperature-force cycles in bar 2 for elastic-viscoplastic (1), viscoelastic viscoplastic (2), and viscoelastic (3) response of the bar.

a model with a viscoelastic viscoplastic bar defined by the equation $\dot{\epsilon} = \dot{S}/AE + S/A\lambda_1 + (S-\sigma_0 A)/A\lambda_2$, where σ_0 decreases with temperature. The comparison of the various $S-T$ diagrams drawn for two temperature levels, as well as for $\lambda_1 \to \infty$ (elastic viscoplastic bar) illustrates the reduction of the thermal stresses, particularly those due to temperature reversal, that is caused by the existence of a viscous component (creep) below the yield stress. When the rate of temperature change is sufficiently slow or the relaxation time very short the yield stress may not be attained at all and the viscoplastic component completely eliminated.

Differences in the rates of temperature application and reversal produce widely different stress conditions; Fig. 8 illustrates this fact by representing for a viscoelastic-plastic bar with temperature-dependent yield stress (the effect of the viscosity above the yield stress

[6] G. H. Heimerl and P. J. Hughes, "Structural efficiency of various aluminum, titanium and steel alloys at elevated temperatures," Natl. Advisory Comm. Aeronaut. Tech. Notes, No. 2975, Washington, 1953.

which, in general, is quite insignificant in comparison to the plasticity, has been neglected) the thermal forces produced by (1) rapid heating and slow cooling, (2) slow heating and rapid cooling, (3) rapid heating and cooling, (4) slow heating and cooling in comparison to the thermal forces in the elastic and in the elastic-plastic bar with constant yield stress. The diagrams have been drawn on the basis of the assumption that the temperature is reversed immediately after T_0 has been attained, and the subsequent temperature cycles are also applied without intermission. If either the upper or the reference temperature levels are sustained for certain periods before reversal or reapplication of the temperature, the applied or the residual temperature forces will partly relax, and the force-temperature diagrams change accordingly. Such relaxation will evidently be much faster at the maximum than at the reference temperature, with the result that the level of residual forces will be further raised.

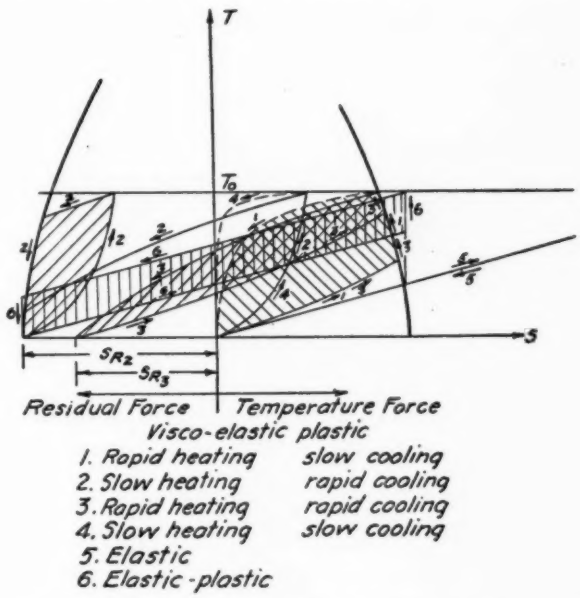


FIG. 8. Temperature-force cycles in bar 2 for viscoelastic-plastic response of the bar and various rates of linear temperature change.

CONCLUSION

From the comparison of the different $S-T$ diagrams for the structural model it is rather obvious that the results of elastic thermal stress analysis are unrelated to the thermal stresses occurring in real materials at elevated temperatures; even the introduction of time- and temperature-independent plasticity, while providing a somewhat better approximation, falls still quite short of reproducing real conditions. These are governed essentially by the temperature-dependent viscosity below the yield stress as well as, to a lesser degree, by the temperature dependence of the yield stress. It appears, therefore, that unless a specific solution of the 3-dimensional thermal stress problem for the particular rheological material exists, a rough estimate of the intensity of the thermal stresses associated with various temperature cycles on the basis of a simple rheological model will provide a better estimate of the thermal stresses, and be a more reliable guide for design than an elaborate elastic analysis.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 9 SEPTEMBER, 1954

Flow of Melts in "Crosshead"-Slit Dies; Criteria for Die Design

JAMES F. CARLEY
Polychemicals Department, E. I. du Pont de Nemours and Company, Inc., Wilmington, Delaware
(Received April 12, 1954)

"Crosshead"-slit dies are widely used for the extrusion of polymer sheeting, flat film, and blown film. In this paper we analyze the laminar flow of liquids in such dies, deriving an equation relating pressure to position along the slit for materials that obey the "power law." The equation yields an index of uniformity (UI), which is the ratio of the rate of extrusion at the far end to that at the feed end, given by

$$UI=\left\{1-\frac{[(1+0.05n)n\alpha L]^{(n+1)/n}}{(1+0.05n)(n+1)}\right\}^{n},$$

where n = the exponent of the power law, L = the length of the die from feed end to far end, and α = a parameter depending on the other die dimensions and the exponent n, having the unit $(\text{in})^{-1}$. An illustrated procedure is given for designing dies that will make film whose caliper is satisfactorily uniform over its whole width. We show that the UI for center-fed dies is always greater than for end-fed dies of the same length.

INTRODUCTION

"CROSSHEAD"–SLIT dies are widely used for the extrusion of polymer sheeting, flat film, and blown film. A long-standing problem connected with such dies is the problem of getting uniform rates of extrusion over the whole length of the die. A difficult part of this problem is the accurate adjustment of the slit depth; another part is the maintenance of uniform temperature over the whole die. These two obstacles are mostly mechanical and can be overcome by taking sufficient pains. Even when they are, there is still a uniformity problem because the pressure of the melt is high at the feed end of the die and low at points that are far from the feed end. This can be better explained with the help of Fig. 1, which is a schematic drawing of a simple crosshead-slit die in which the feed enters at one end. Reduced to barest essentials, it is a pipe open at one end and plugged at the other, with a longitudinal cut along its whole length. As the polymer enters the open end, some of it is squeezed out of that end of the slit, while most of it continues along the pipe. The flow along the pipe is resisted by the drag of the melt on the pipe wall. Some of the available head is spent in overcoming this resistance, so the pressure drops as the melt moves along the pipe. The pressure outside the die lips, however, stays the same. The further we go from the entrance, therefore, the smaller is the pressure drop available to force the melt through the slit. The local rate of extrusion at the far end of the slit must be less than it is at the feed entrance. If the sheet is drawn off at the same speed everywhere, then it must become thinner at the far end. This situation of unequal extrusion rates raises several questions: How bad is the inequality? How is it affected by changes in die dimensions and polymer properties? What can be done to improve uniformity? The rest of this paper is devoted to finding answers to these questions.

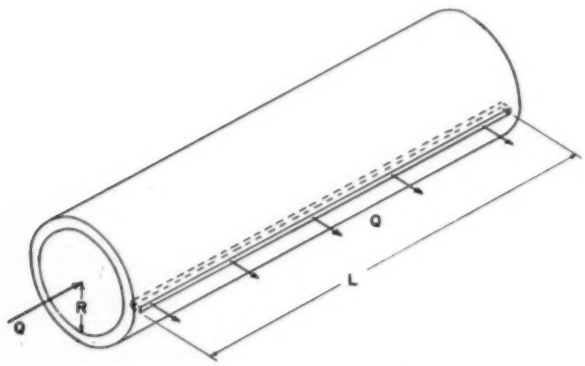


FIG. 1. Crosshead-slit die.

Before beginning the derivation of the equations describing the behavior of melts in such dies, let us see what we can learn by inspection. It is clear that if the pipe diameter were very large in comparison to the slit thickness, the pressure drop from entrance to end would be a very small part of the initial pressure, and the uniformity would be good. This hints to us that the pressure at which the melt enters the die is not important of itself; rather what matters is the division of this available head between the channel and the lips. Also, we can see that, if other factors are kept the same, the longer the slit, the more difficult it will be to attain satisfactory uniformity. With these expectations in mind we can get on with the development.

The steady flow of a liquid along a circular pipe is called one-dimensional flow because the local velocity of the fluid varies in only one direction—radially. Similarly, in a long, shallow slit, since the ends may be neglected, the velocity effectively varies only in the direction of the slit depth and is constant across the length. The cases we shall consider here all involve the combination of these two types of one-dimensional flow. Flow through a narrow annulus fits into either class but is most easily handled by considering the annulus to be a slit, without ends, that has been bent into a circle.

DERIVATION OF EQUATIONS

One-dimensional flow of a liquid may be described by the equation

$$dv/dx=g(s,f), \qquad (1)$$

where v = the velocity at any point x and x = a point on the axis perpendicular to the direction of flow. The x axis is the direction of the velocity gradient dv/dx. $g(s,f)$ = an unspecified function of the shear stress s at x, and the nature of the fluid. In Newtonian liquids $g(s,f) = s/\mu$, where μ = the viscosity. The term viscosity, in the rigorous sense at least, applies only to Newtonian fluids. A great deal of confusion has come from applying the term to non-Newtonians.

If the liquid does not slip at the channel wall, the flow is further described by the boundary condition

$$v(x_w)=0, \tag{2}$$

where x_w = the value of x at the wall. The shear stress is dependent on the geometry and in general can be expressed in terms of the x coordinate and the pressure gradient. Equation (1) may be integrated with respect to x to give the local velocity

$$v(x)=\int g(s,f)dx+C. \tag{3}$$

C may be evaluated from Eq. (2). If the liquid is incompressible, the volume rate of flow through the channel is found by integrating the velocity over the cross section:

$$Q=\int_{x_0}^{x_w} v(x)dA. \tag{4}$$

This gives the discharge in terms of the pressure gradient, the geometry of the channel and the properties of the fluid.

Before going on I would like to introduce a specific stress function, the so-called power law. The operations we will do can just as well be done with other functions. But they cannot be done at all, only indicated, unless a specific function is chosen. The power function is a good choice since it satisfactorily represents the behavior of many polymer melts and solutions. It also includes the Newtonian case. Thus Eq. (1) becomes

$$dv/dx=f_0(s/s_0)^n, \tag{1a}$$

where s_0 = the stress corresponding to the "standard state," f_0 = the shear rate when $s=s_0$, and n = the power or exponent, usually between 1 and 4, which gives the function its name. Both n and f_0 are characteristic of the material and f_0 is dependent on temperature. I have stated the power law in this form to avoid the dimensional objection to which the usual form of the law is open.[1] The standard stress s_0 may be any stress we like; its units are chosen to be convenient for the uses to which the equations are to be put. *The units of the other terms must be consistent with those of* f_0. I have chosen a standard stress of 1 psi. To avoid a confusion of terms and gain generality, I call f_0 the "generalized fluidity" of the material. When $n=1$, which is the Newtonian case, generalized fluidity is synonymous with fluidity or reciprocal viscosity.

When a material obeying (1a) flow isothermally through a circular channel, the shear rate is given by

$$\frac{dv}{dr}=f_0\left(\frac{s}{s_0}\right)^n=f_0\left(\frac{r}{2}\frac{dP}{dz}\right)^n=\frac{f_0 r^n}{2^n}\left(\frac{dP}{dz}\right)^n, \tag{1b}$$

where $s_0=1$ psi, r = the radius from the center, $0\le r\le R$, in., R = the radius of the channel, in., and dP/dz = the pressure gradient, psi/in. The direction of the pressure gradient is opposite that of the flow; *throughout this paper the positive z direction is taken as the direction of the pressure gradient.* The origin of the z axis is therefore the discharge end of any channel being considered. This guarantees that the pressure gradient is always positive and keeps us from having negative numbers raised to nonintegral powers. Integrating and substituting the boundary condition $v(R)=0$.

$$v(r)=-\frac{f_0}{2^n(n+1)}\left(\frac{dP}{dz}\right)^n[R^{n+1}-r^{n+1}]. \tag{3b}$$

The discharge is given by Eq. (4):

$$Q=\int_0^R v(r)2\pi r dr$$

$$Q=-\frac{\pi f_0 R^{n+3}}{2^n(n+3)}\left(\frac{dP}{dz}\right)^n. \tag{4b}$$

Similarly, if the channel is a narrow slit, the discharge is

$$Q=-\frac{f_0 L h^{n+2}}{2^{n+1}(n+2)}\left(\frac{dP}{dz}\right)^n, \tag{4c}$$

where L = the length of the slit, in inches, and h = the depth of the slit, in inches. Recalling that s_0^n is present in the denominators of Eqs. (4b) and (4c), we see that the units of Q are in.^{3}/sec.

It will help now to look at Fig. 2, which shows a small slice cut out of a crosshead-slit die. Over the short length dz of this slice the pressure is constant, but the flow along the axis is diminished by the amount dQ that flows out the slit. Applying Eq. (4b) at the two points z and $z+dz$ and then subtracting, we get

$$Q+dQ-Q=-\frac{\pi f_0 R^{n+3}}{2^n(n+3)}\left[\left(\frac{dP}{dz}\right)^n+\frac{d}{dz}\left(\frac{dP}{dz}\right)^n dz\right]$$

$$+\frac{\pi f_0 R^{n+3}}{2^n(n+3)}\left(\frac{dP}{dz}\right)^n.$$

[1] M. Reiner, *Deformation and Flow* (H. K. Lewis & Company, London, 1949), p. 102, discusses this and other forms of the power law; also presents Eq. (4b) of this paper.

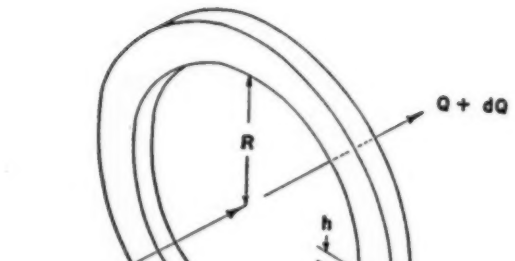


FIG. 2. Z element of crosshead-slit die.

Taking the indicated derivative and dividing by dz,

$$\frac{dQ}{dz}=-\frac{\pi n f_0 R^{n+3}}{2^n(n+3)}\left(\frac{dP}{dz}\right)^{n-1}\frac{d^2P}{dz^2}. \tag{5}$$

In a crosshead die the pressure gradient in the slit is directed inward while the pressure gradient in the pipe is directed axially toward the entrance. This means that the z directions in Eqs. (4b) and (4c) are not the same but are perpendicular. Since we are not concerned with local variations in pressure across the lips of the die, and since we are dealing only with dies having parallel lips, we may replace dP/dz in Eq. (4c) with P/t. Hereafter, the z direction will be the direction of the axis of the cylindrical channel. The increment of discharge through a length dz of the slit is therefore

$$dQ=-\frac{f_0 h^{n+2}}{2^{n+1}(n+2)}\left(\frac{P}{t}\right)^n dz$$

and

$$\frac{dQ}{dz}=-\frac{f_0 h^{n+2}}{2^{n+1}(n+2)}\left(\frac{P}{t}\right)^n. \tag{6}$$

Combining (5) and (6), we have

$$\frac{d^2P}{dz^2}\left(\frac{dP}{dz}\right)^{n-1}-\alpha^{n+1}P^n=0, \tag{7}$$

where

$$\alpha^{n+1}=\frac{(n+3)h^{n+2}}{2\pi n(n+2)t^n R^{n+3}}. \tag{8}$$

Note that α has the dimension of reciprocal length for all values of n.

Equation (7) is the fundamental equation describing the behavior in crosshead-slit dies of fluids obeying the power law. By the same procedure used here similar equations can be derived for other basic types of behavior. Equation (7) is a second-order differential equation of the nth degree. It can be reduced[2] to a first-order equation of the $(n+1)$th degree by multiplying through by dP/dz:

$$\frac{d^2P}{dz^2}\left(\frac{dP}{dz}\right)^n-\alpha^{n+1}\left(\frac{dP}{dz}\right)P^n=0.$$

In this form we see that

$$\frac{d^2P}{dz^2}\left(\frac{dP}{dz}\right)^n=\frac{1}{n+1}\frac{d}{dz}\left(\frac{dP}{dz}\right)^{n+1}$$

and

$$\left(\frac{dP}{dz}\right)P^n=\frac{1}{n+1}\frac{d}{dz}P^{n+1},$$

while

$$0=\frac{d}{dz}C.$$

Integrating all terms, we have

$$\frac{1}{n+1}\left(\frac{dP}{dz}\right)^{n+1}-\frac{\alpha^{n+1}}{n+1}P^{n+1}=C. \tag{9}$$

Let us refer, for a moment, to Eq. (4b). At $z=0$, the discharge is zero, since the die is closed at that end. The last element dQ escapes from the slit just beyond $z=0$. Thus, at that point dP/dz must be zero, by Eq. (4b). That is,

$$\left(\frac{dP}{dz}\right)_{z=0}=0. \tag{10}$$

If this boundary condition is applied to Eq. (9) we have

$$C=-\frac{\alpha^{n+1}P_f^{n+1}}{n+1}. \tag{11}$$

Separating the variables of Eq. (9), we have

$$\frac{dP}{[(n+1)C+\alpha^{n+1}P^{n+1}]^{1/n+1}}=dz. \tag{12}$$

Now, let P_i = the entering melt pressure. We define two new quantities ν and ϵ by the equations

$$P=(1-\nu)P_i$$

$$P_f=(1-\epsilon)P_i \quad (\text{where } \epsilon=\nu_f=\nu \text{ for } z=0).$$

If these substitutions and Eq. (11) are combined with Eq. (12), we have

$$dP=-P_i d\nu$$

$$(n+1)C=-\alpha^{n+1}(1-\epsilon)^{n+1}P_i^{n+1}$$

$$\alpha^{n+1}P^{n+1}=\alpha^{n+1}(1-\nu)^{n+1}P_i^{n+1}.$$

[2] E. Kamke, *Differentialgleichungen, Losungsmethoden und Losungen* (Akademische Verlagsgesellschaft, Leipzig, 1943), second edition, p. 596.

Thus,

$$\frac{dv}{[(1-v)^{n+1}-(1-\epsilon)^{n+1}]^{1/n+1}}=-\alpha dz. \quad (13)$$

The left side of Eq. (13) can be integrated in closed form only for certain special values of n such as $n=1$, 1.5, 2, etc. In practice, we must deal with materials for which n may take on values like 2.4, 3.2, and so on. The change of variable from P to v facilitates making an approximation to the original function which will yield closed solutions for all n. First we should note that the range of interest for ϵ, the fractional loss of pressure from one end of the slit to the other, is $0<\epsilon<0.1$, or thereabouts. If $\epsilon=0.1$, the extrusion rate at the far end will normally be 80 percent or less of that at the entrance end. This is poor uniformity and in designing dies we shall strive to make ϵ much less than 0.1. Note also that $0\leq v\leq\epsilon$ in any one die.

The binomial expansion of $(1-v)^{n+1}$ is

$$(1-v)^{n+1}=1-(n+1)v+\frac{(n+1)n}{2!}v^2-\cdots \quad (14)$$

also

$$(1-\epsilon)^{n+1}=1-(n+1)\epsilon+\frac{(n+1)n}{2!}\epsilon^2-\cdots. \quad (15)$$

If we subtract Eq. (15) from Eq. (14) we have

$$(1-v)^{n+1}-(1-\epsilon)^{n+1}=(n+1)(\epsilon-v)-\frac{(n+1)n}{2!}(\epsilon^2-v^2)+\cdots. \quad (16)$$

This expression is largest when $z=L$, where $v=0$ and $\epsilon\leq 0.1$. If only the first term of the series is used, the error is very nearly $-5n$ percent of the true value. With an error of less than 1 percent we can write

$$(1-v)^{n+1}-(1-\epsilon)^{n+1}=(1+0.05n)(n+1)(\epsilon-v). \quad (17)$$

This is clearly linear in v and with the aid of the substitution

$$u=(1+0.05n)(n+1)(\epsilon-v) \qquad dv=-\frac{du}{(1+0.05n)(n+1)}. \quad (18)$$

Equation (13) reduces to the easily integrated form

$$\int_{u_0}^{u_L}\frac{du}{u^{1/(n+1)}}=+(1+0.05n)(n+1)\alpha\int_0^L dz, \quad (19)$$

where

$$u_0=(1+0.05n)(n+1)(\epsilon-\epsilon)=0$$
$$u_L=(1+0.05n)(n+1)(\epsilon).$$

Integrating and solving for $1-\epsilon$, we have

$$1-\epsilon=\frac{P_f}{P_i}=1-\frac{[(1+0.05n)n\alpha L]^{(n+1)/n}}{(1+0.05n)(n+1)}. \quad (20)$$

The distribution of pressure along the slit is of some interest. It is found by integrating Eq. (19) between slightly different limits:

$$\int_{u_0}^{u}\frac{du}{u^{1/(n+1)}}=+(1+0.05n)(n+1)\alpha\int_0^z dz,$$

where

$$u=(1+0.05n)(n+1)(\epsilon-v).$$

This gives

$$P=P_i\left\{1-\frac{(1+0.05n)^{1/n}(n\alpha)^{(n+1)/n}}{n+1}\times[L^{(n+1)/n}-(L-z)^{(n+1)/n}]\right\}. \quad (21)$$

Equation (20) holds the greatest practical interest since it leads, along with Eq. (6), to an index of uniformity. Referring back to (6), we see that the ratio of the discharge rates at opposite ends of the die, dQ_f/dQ_i is just

$$\frac{dQ_f}{dQ_i}=\left(\frac{P_f}{P_i}\right)^n=(1-\epsilon)^n. \quad (22)$$

$(1-\epsilon)^n$, therefore, is a direct measure of uniformity and we have named it the *uniformity index* or *UI*.

We have a partial check on the error incurred in Eq. (20) by the linearizing approximation. When $n=1$, the Newtonian case, the original differential Eq. (7) simplifies to the familiar linear differential equation

$$\frac{d^2P}{dz^2}-\alpha^2P=0. \quad (23)$$

The exact solution of this equation is

$$1-\epsilon=\frac{P_f}{P_i}=\frac{1}{\cosh\alpha L}=\operatorname{sech}\alpha L. \quad (24)$$

If $n=1$ is substituted into Eq. (20), we have

$$1-\epsilon=\frac{P_f}{P_i}=1-\frac{(1.05\alpha L)^2}{2.10}. \quad (25)$$

The right side of (25) is just the first two terms of the expansion of $\operatorname{sech}\alpha L$, slightly altered. Both equations contain only the dimensionless parameter αL. This in turn depends only on the dimensions of the die. Thus, as we had suspected, the uniformity of extrusion is independent of viscosity and the entering pressure. In non-Newtonians, of course, the characteristic exponent n plays an important part. Note that the

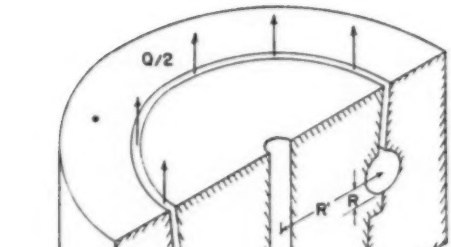


FIG. 3. Die for blown tubing (half-section).

generalized fluidity f_0 has dropped out, however. As a check on accuracy, then, we calculated $1-\epsilon$ from both Eqs. (24) and (25) for several values of αL. The results are in Table I. Thus, in the range $0 \leq \epsilon \leq 0.1$, the error is 1 percent or less. We can expect that for $n=3$, say, it will be slightly larger. These errors are well within the errors imposed by other limitations of the problem (e.g., the closeness with which real melts obey the power law; ignoring end effects and heat effects).

What happens to uniformity as we switch from a Newtonian melt to increasingly non-Newtonian ones? This amounts to raising n while keeping the die dimensions constant. The parameter α was defined by Eq. (8). From this equation we can see that if h, R and t are kept constant α will depend only on n.

To show how strongly the uniformity depends on the exponent n, we have calculated the uniformity indexes for materials of $n=1, 2, 3, 4$, flowing through a film die. The die is 48 inches long, center-fed, so the length for our calculations is 24 inches. The other dimensions are $R=0.75$, $h=0.01$, $t=0.50$ inch. For each n we calculated αL and UI from Eqs. (8) and (20). Here are the results:

n	αL	UI
1	0.0277	1.000
2	0.0614	0.970
3	0.0986	0.852
4	0.1348	0.665

For materials having n's in the range 1–2.5, this die will be fairly satisfactory, at least by the present standards of the film industry. If $n=3$, undrawn film from this die would have 15-percent variation in caliper from center to edges. For $n=3.2$ and higher the variation in caliper would be much too high.

DISCUSSION

Uniformity of caliper is important to good winding, to later handling in making bags, etc., and in many of the consumer applications. A certain minimum thickness is needed in a bag, for example, which is to hold a dozen oranges. Any heavier thickness is a waste of polymer and needlessly raises the cost.

To meet the minimum caliper specification with a UI of 0.85 requires that about seven percent more polymer be extruded per yard of film than is actually needed. Production stoppages in fabricating equipment can be expensive in terms of film and labor wasted. An important reason for using high draw ratios in making polyethylene film is to get the caliper control that present dies fail to give. And high draw has disadvantages of its own.

Granting that uniformity at the die is worth shooting for, how do we get it? Generally we are stuck with length L and the thickness h, but we are at liberty to play with the channel radius R and the land width of the lips t. A glance at Eq. (8) reveals that the parameter α is much more sensitive to changes in R than in t, since

$$\alpha \sim R^{-(n+3)/(n+1)} t^{-n(n+1)}. \tag{26}$$

If in the 48-inch die we were considering R were doubled then for $n=3$, the $UI=0.967$ instead of 0.852. If t were doubled instead, the $UI=0.933$. Increasing t, however, will require a rise in the pressure level in order to maintain the desired output. Increasing R, on the other hand, will very slightly reduce the required pressure. There are disadvantages: a bigger radius means longer holdup in the hot die and less rigidity.

From Eq. (20) we can see that the greater the length L, the poorer will be the uniformity. For this reason a *center-fed die is always preferable to an end-fed die of the same total length*. If the 48-inch die were end-fed the UI would be 0.693 instead of 0.852, a very serious loss of uniformity.

Our analysis also applies to certain dies for making blown tubing. A half-section of such a die is sketched in Fig. 3. Here the melt enters a doughnut-shaped channel and passes out through the annular slit which runs around the top of the torus. This amounts to a center-fed flat-sheeting die which has been bent around a center and which has no ends. If the ratio of the radii R/R' is 0.2 or less, the foregoing formulas apply with little error. For the die length L we must substitute $\pi R'$; nothing else is changed.

The expression for the uniformity index, together with the equation for the discharge through a narrow slit, provide a rational basis for the design of crosshead dies. Given the rate and uniformity desired, the avail-

TABLE I.

	$1-\epsilon = dQ_f/dQ_i = UI$		
αL	Exact	Approximate	Percent relative error
0.001	1.000	1.000	0.0
0.01	0.999	0.999	0.0
0.1	0.995	0.995	0.0
0.4	0.925	0.916	−1.0
0.5	0.887	0.869	−2.0

able melt pressure, the slit length and depth, and the generalized fluidity of the melt at the extrusion temperature, the remaining die dimensions can be determined. The practical implications of these results and their employment in an ordered procedure for designing dies will form the subject of another paper now in preparation.

Throughout this paper we have assumed that the temperature of the melt is the same everywhere in the die. Practical achievement of this condition in large dies is no mean feat, as some of you know. I don't wish to begin a discussion of ways and means of getting uniform temperature, except to say that we believe it can be done.

To sum up, we have derived a differential equation describing the laminar flow of liquids obeying the power law in crosshead-slit dies. The solution of this equation yielded a measure of uniformity of extrusion from the extreme ends of the die, the uniformity index. We have shown that we can get uniformity of film caliper by choosing the proper die dimensions, thus minimizing the need for arduous and costly point-to-point adjustment of the slit opening. Where possible, film dies should be center-fed rather than end-fed. Finally, the way has been pointed to the use of other stress-*versus*-shear rate relationships in designing film dies.

DEFINITIONS OF PRINCIPAL TERMS

f_0 = the "generalized fluidity" of a material that behaves according to the power law, sec^{-1}. It is equal to the shear rate when the shear stress is equal to the shear stress of the standard state s_0. In this paper s_0 has been taken as 1 psi.

h = the depth of the slit in a slit die, in. If the sheet were undeformed after leaving the die, h would be the sheet thickness.

n = the exponent in the power law.

L = the length of a slit die from the far end of the slit to the feed end, in. If the sheet were undeformed after leaving the die, this would be the sheet width.

P = the pressure of the melt behind the slit at any point z on its length, psi (gauge).

P_i = the pressure of the melt at the feed port, i.e., at $z=L$.

P_f = the pressure of the melt at the far end of the slit, i.e., at $z=0$.

Q = the volume rate of flow along the supply channel behind the slit at any point z, $in.^3/sec$.

dQ_i = the discharge from the increment of length adjacent to the feed port, corresponding to P_i.

dQ_f = the discharge from the increment of length adjacent to the far end of the slit, corresponding to P_f.

r = the radial distance of a point from the axis of the supply channel, in.

R = the radius of the supply channel (= the maximum value of r).

s = the shear stress at any point in the melt, psi.

t = the width of the die slit, i.e., the dimension in the direction of flow through the slit, in.

UI = the uniformity index = dQ_f/dQ_i, dimensionless.

v = the linear velocity of flow at any point in the melt, in./sec.

x = the direction of the velocity gradient, perpendicular to the flow and to the pressure gradient, in.

z = the direction of the pressure gradient in any channel, in. In a crosshead sheeting die this is the direction along the axis of the supply channel. z increases as one goes from the far end, where $z=0$, to the feed end, where $z=L$ (or $L/2$ in some dies).

α = a parameter important in the characterization of crosshead dies, $in.^{-1}$ It is defined by Eq. (8).

ϵ = the value of v at $z=0$.

ν = the fractional loss in pressure at any point z, i.e., $\nu = 1-(P/P_i)$.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 9 SEPTEMBER, 1954

Rotational Plastometry Applied to Molten Polyethylene

F. D. DEXTER
Development Laboratories Bakelite Company, A Division of Union Carbide and Carbon Corporation, Bound Brook, New Jersey
(Received April 12, 1954)

A concentric cylinder rotational plastometer which automatically plots the shear strain *vs* time curve for each of a number of shear stresses is described. The flow behavior of molten polyethylene resins, measured with this instrument from 130°C to 210°C, is discussed. It is shown that, at each shear stress, the shear strain *vs* time curve for these resins is describable in terms of a viscosity, an elastic shear strain and a retardation time. As the shear stress is raised both the viscosity and retardation time decrease, while the elastic shear strain increases. The shear stress dependency of the viscosity and elastic shear strain are shown to be related and in a manner which suggests that the viscosity change is the result of an entropy effect.

I. INTRODUCTION

THE fundamental problem in the rheology of molten thermoplastics is that of separating and measuring the various components of the flow. It is generally accepted that the deformation at any given time results from the simultaneous response of both viscous and elastic flow components. This gives rise to a nonlinear deformation-time curve. Furthermore, it is usually conceded that these components do not respond to the applied shear stress linearly, except at very small shear stresses. Thus, in order to determine, quantitatively, the nature of this nonlinear dependency of flow on time and shear stress, test conditions which permit the measurement of the deformation as a time function, under conditions of known shear stress, are required.

A parallel plate type plastometer can be used to measure deformation-time properties,[1] but in order to solve the equations associated with this process, it is necessary to assume linearity of response with respect to applied stress. Furthermore, the actual stress is a function of the plate separation, which decreases with time, and is thus a time function which varies according to the nature of properties being sought. Consequently, the test conditions must be limited to shear stress regions where linear response can reasonably be expected and only time dependent effects can be measured.

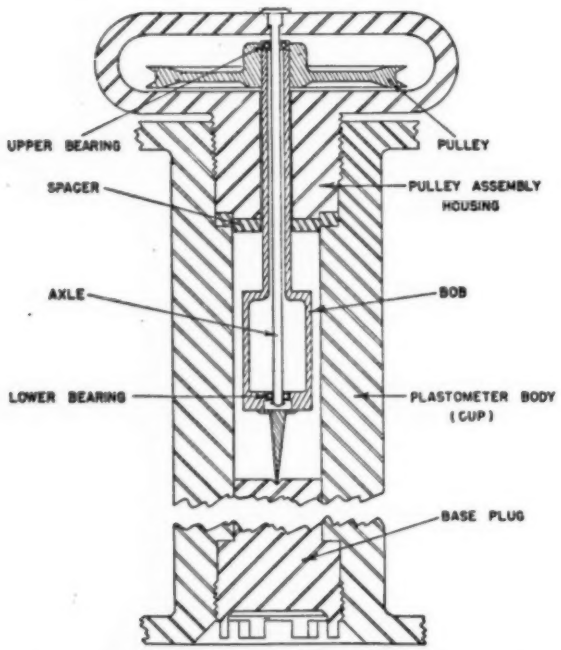


FIG. 1. Cross-section drawing of rotational plastometer.

Extrusion plastometer techniques can also be used to measure flow behavior of molten thermoplastics.[2,3] While the resulting data are practically useful the difficulties in reducing them to fundamentally significant quantities are serious. The problem here is twofold. First the shear stress, at any constant extrusion pressure, varies from zero at the center of the extrusion capillary to some maximum value at the capillary wall. This produces a velocity front whose shape is determined by material properties whose values are being sought. Second, time dependent properties exert an influence which depends on the time material is under the influence of the shearing stress[4] (i.e., time in the capillary). This time will vary from some minimum at the center of the capillary to infinity at the capillary wall. Thus, the nature of these time dependent properties will also alter the shape of the velocity front and will, no doubt, produce anomalous entrance effects. The net effect is one in which the observed phenomena are the result of both time and stress dependent effects whose relative importance can only be assessed in a general manner.

Concentric cylinder rotational plastometers largely overcome the objections based on nonuniformity of applied stress. However, it is common practice to observe only the steady-state rate of shear (rotation), resulting from a known shear stress, and thus time dependent effects remain unknown. Only fragmentary interest has been shown in measuring the strain as a function of time.[5,6] In order to do this some form of a recording mechanism is required. It is the purpose of this paper to describe a recording concentric cylinder

[1] G. J. Dienes, J. Coll. Sci. **2**, 131 (1947).

[2] H. K. Nason, J. Appl. Phys. **16**, 338 (1945).
[3] R. S. Spencer and R. E. Dillon, J. Coll. Sci. **4**, 241 (1949).
[4] F. D. Dexter and G. J. Dienes, J. Coll. Sci. **5**, 228 (1950).
[5] W. Braunbek, Z. Physik **57**, 501 (1929).
[6] R. Buchdahl, J. Coll. Sci. **3**, 87 (1948).

plastometer and to discuss the flow behavior of molten polyethylene as determined using this instrument.

II. APPARATUS AND EXPERIMENTAL PROCEDURE

The rotational plastometer constructed for this work was of generally conventional design, consisting of a rotable cylinder (bob) mounted concentrically in a cylindrical cavity (cup). Temperature was controlled by means of a thermostated bath surrounding the cup. A drawing of the assembled apparatus is shown in Fig. 1. The critical dimensions are

Length of bob	5.08 cm
Radius of bob	1.662 cm
Radius of cup	1.916 cm
Radius of pulley	5.16 cm.

Shear stress is applied to material contained between the cup and the bob by dead loading a cord (instrument cable) wrapped around the pulley on the bob. The cord system is arranged so as to produce a couple thus avoiding any bending moment on the bob. The shear stress and resulting shear rate are both calculable on the basis of well-known equations.[7,8] Application of these show that

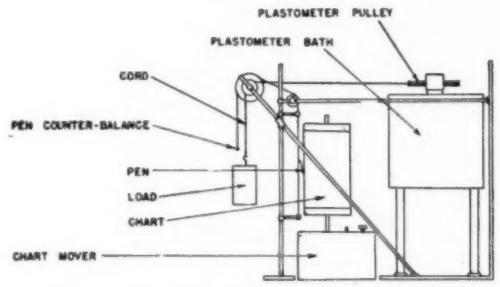


FIG. 2. Diagram of recording system for rotational plastometer.

$$\text{mean shearing stress (dynes/cm}^2\text{)}, S = 52.8W \quad (1)$$

and

$$\text{mean shearing rate (sec}^{-1}\text{)}\ \dot{\gamma} = 44.0 \times \text{rev/sec}, \quad (2)$$

where W = load on cord system in grams. From Eq. (2) it can be easily shown that

$$\text{mean shear strain}, \gamma = 0.122 \times \theta, \quad (3)$$

where θ = angular rotation of bob in degrees. This is equivalent to about 8 degrees rotation for unit shear strain. In all of this work end effects have been ignored and it will be seen that the error so introduced is small.

A recording system was designed to plot the angular displacement of the bob, which is directly proportional to the shear strain [Eq. (3)], against time. The time axis was obtained through use of a chart mover* having 22 available chart speeds ranging from 0.00125 to 24.33 cm/sec. A pen, driven by the rotation of the bob, traced the displacement curve on this chart. This arrangement is illustrated in Fig. 2. With the system illustrated 2.38 cm chart travel corresponds to unit shear strain.

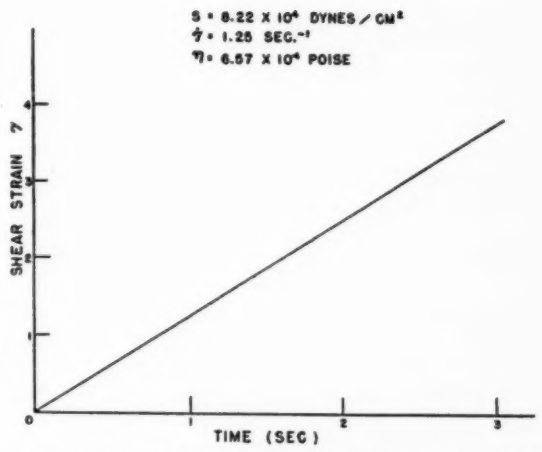


FIG. 3. Typical shear strain *vs* time curve for phenol-formaldehyde resin at 120°C.

After adjusting the thermostated bath to the desired temperature material was placed in the cup and subsequently tamped to remove air bubbles. The bob, which had been kept hot in an oven, was then inserted and the cord system assembled. A stop was placed in the pulley on the bob and the desired load attached to the cord. The chart mover was then started, and after it had been allowed to rotate a revolution or so to trace a baseline, the stop was removed. At the termination of the chart travel the stop was reinserted. Sufficient time was allowed to elapse between tests to assure complete relaxation of the material.

III. RESULTS

A. Check Tests with Novolak Resin

Initial tests were run on a phenol-formaldehyde novolak resin. The purpose of these was to check the calculated instrumental calibration and to ascertain that the recording system functioned properly. Since resins of this type exhibit only a viscous flow component[9] they are very convenient for work of this nature. A typical curve showing the shear strain as a function of time, determined using the recording system, is given in Fig. 3. This curve exhibits the expected linearity with no evidence of any elastic deformation. Thus, it can be stated that curves showing evidence of elastic like phenomena are the result of material properties and not instrumental peculiarities.

The slope of a shear strain *vs* time curve defines a shear rate and the load on the cord system determines the shear stress. By a number of determinations at various loads, shear rate *vs* shear stress data were obtained for the novolak resin. This same resin was also

[7] J. M. Burgers, *Second Report on Viscosity and Plasticity* (North Holland Publishing Company, Amsterdam, 1938).
[8] E. K. Fischer and C. H. Lindsey, J. Coll. Sci. 3, 111 (1948).
* Gorrell and Gorrell "Monodrum" Kymograph.
[9] G. J. Dienes, J. Coll. Sci. 4, 257 (1949).

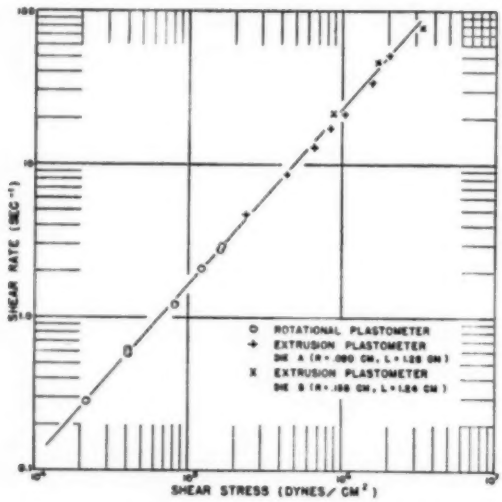


FIG. 4. Shear rate *vs* shear stress curve for phenol-formaldehyde resin at 120°C.

tested in an extrusion plastometer, patterned after Nason's high-pressure rheometer,[2] from which shear stress and shear rate data were likewise computed. The results from both the rotational and extrusion plastometer are shown in Fig. 4; the adjustment to the extrusion shear rate, suggested by Rabinowitsch,[10] proved to be within test error and was not made. It is seen that all these data define a single curve. The calculated constants for the rotational plastometer are, therefore, judged to be correct within test error.

B. Polyethylene

Molten polyethylene has been shown to exhibit both viscous and elastic flow components.[1] The curve shown in Fig. 5, which was obtained with the rotational plastometer and recording system on a sample designated resin *A*, is typical of viscoelastic materials. The purpose of this investigation was to determine how the characteristics of such curves vary as the applied shear stress is changed. In order to do this it is necessary to define the curves mathematically. The most rigorous definition probably consists of the equation which describes the deformation of a number of Voight elements (parallel spring and dashpot) in series.

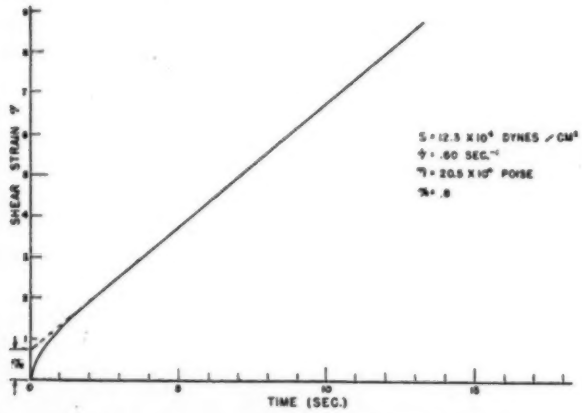


FIG. 5. Typical shear strain *vs* time curve for polyethylene resin *A*.

[10] B. Rabinowitsch, Z. physik. Chem. **A145**, 1 (1929).

TABLE I. Newtonian viscosity for three polyethylene resins.

Resin	Newtonian viscosity at 170C (poise)
A	7.5×10^5
B	2.2×10^5
C	1.2×10^5

$$\gamma=\sum_{i=1}^{i=n}\gamma_i[1-\exp(-t/\tau_i)], \tag{4}$$

where γ_i = maximum shear strain of ith element; τ_i = retardation time of ith element; t = time. However, from a curve of the type shown in Fig. 5 only a limited number of components can be defined. The straight line portion, which can be described in terms of viscosity, evidently results from all elements having retardation times substantially longer than the test time interval. The remainder of the elements are lumped into a "total elastic shear strain" and a retardation time. This latter most likely possesses some of the properties of a "most prominent relaxation time."[11] The simplified equation, describing the flow at a given shear stress, is

$$\gamma=St/\eta+\gamma_e[1-\exp(-t/\tau_0)], \tag{5}$$

where η = observed viscosity; γ_e = total elastic shear strain; τ_0 = retardation time. The effect of shear stress on the flow will be discussed in terms of the above parameters. Within certain limitation γ_e could be converted to an equivalent modulus but it was found that the strain itself is more convenient to use.

The effect of shear stress on the viscosity term of Eq. (5) is illustrated in Fig. 6 for polyethylene resin *A*. Curves of a generally similar nature were found for two other resins designated *B* and *C*. One point of interest is that, as the shear stress falls, the degree of dependence of viscosity on shear stress decreases. Expanding this idea a little leads to the conclusion that at very low shear stresses the material is Newtonian in its flow behavior. Spencer and Dillon[3] have reported a similar type of behavior for polystyrene resin. Values for this Newtonian viscosity, obtained by an extrapolation technique,[3] are given in Table I for the three polyethylene resins.

TABLE II. Heats of activation for three polyethylene resins.

Resin	ΔH_v, kcal/mole
A	12.90
B	13.95
C	11.90

[11] W. A. Yager, Physics **1**, 434 (1936); K. W. Wagner, Ann. Physik **40**, 817 (1913).

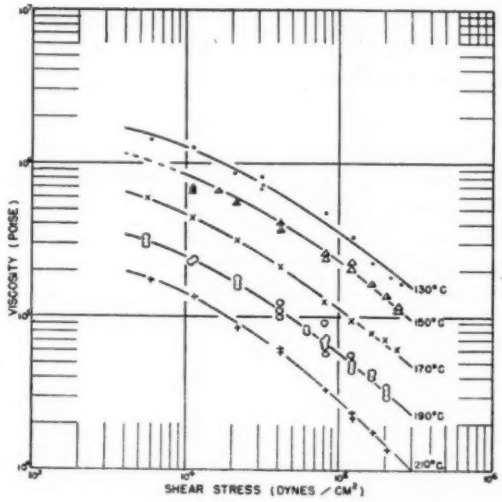


FIG. 6. Viscosity *vs* shear stress curves for polyethylene resin *A*.

Casual inspection of Fig. 6 suggests that the temperature sensitivity of the observed viscosity, at constant shear stress, is independent of the shear stress. An analysis of the data bears this out. The temperature dependency of the viscosity, at any given shear stress, was found to follow the well-known equation,

$$\eta = Ae^{\Delta H_v/RT}, \tag{6}$$

where ΔH_v = heat activation for viscous flow; R = gas constant; A = a shear stress function; T = temperature. The heat of activation was found to be constant over the range of shear stresses employed. Values for the three polyethylene resins are given in Table II.

The elastic shear strain γ_e was found to increase with shear stress and to be independent of temperature. This is illustrated in Fig. 7 showing that, within test error, the elastic shear strain for resin *A* at three temperatures defines a single curve which is linear with respect to the logarithm of the shear stress. A similar situation was found for resins *B* and *C*. This behavior can easily be fitted by the equation

$$\gamma_e = c \ln(S/S_0), \tag{7}$$

where c = slope of the curve and S_0 = stress at zero elastic strain. Values for the constants n and S_0 are shown in Table III. The simultaneous change of both the elastic shear strain and viscosity with shear stress and the fact that the viscosity appears to be predominantly Newtonian in the shear stress region equal to S_0 suggests a connection between the two. Evidence that such a connection does exist and is physically real will be given in a following section.

TABLE III. Elastic shear strain constants for three polyethylene resins.

Resin	c	S_0 (dynes/cm²)
A	0.174	1.0×10^3
B	0.133	1.0×10^3
C	0.291	4.0×10^3

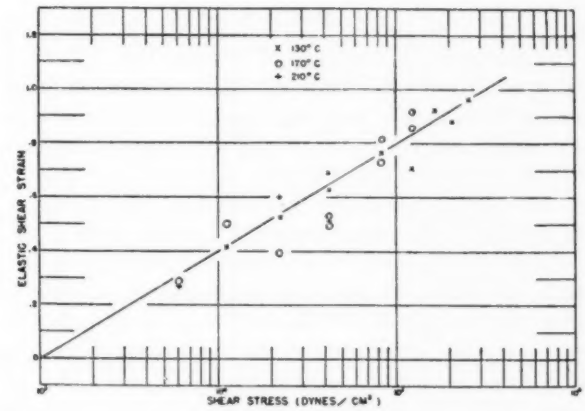


FIG. 7. Elastic shear strain *vs* shear stress curve for polyethylene resin *A*.

The nonlinear portion of the shear strain *vs* time curves, such as is shown in Fig. 5, cannot be described perfectly by a single exponential term. However, the precision of the data do not warrant calculation of more than one retardation time from each curve. Because of this the relation between the elastic shear strain and the retardation time which would be predicted on the basis of a single Voight element is not found. It seems probable that the computed retardation time is most closely associated with a "most prominent relaxation time."[11] Whatever the nature of this value, it was found to fall sharply with increase in either shear stress or temperature. This is illustrated in Fig. 8 for polyethylene resin *A*. Similar curves were found for resins *B* and *C*.

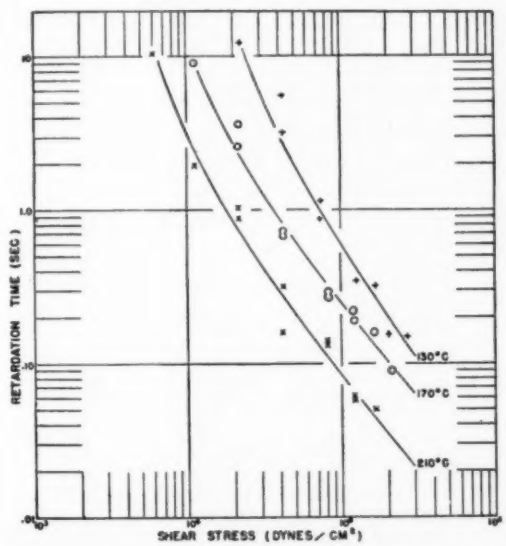


FIG. 8. Retardation time *vs* shear stress curves for polyethylene resin *A*.

IV. ANALYSIS OF DATA

The intent in this section is to show that (1) there exists a real connection between the shear stress dependence of the viscosity and the magnitude of the elastic shear strain and (2) the nature of this dependence supports the view that the change in viscosity is associated with an entropy effect. Before going into this, however, it is in order to examine these data in the light of the theory of Eyring and co-workers.[12] According to this theory the drop in viscosity should be accompanied by a change in the heat of activation,[13] defined according to Eq. (6). In the case of the data in Fig. 6 it can be shown that a drop in viscosity of a factor of 10 should be accompanied by a drop in the heat of activation of around 2 kcal/mole. Actual computations show no significant change in the heat of activation over this range and the statistical variations are too small to absorb an effect of this magnitude.†

An analysis of the viscosity data is simplified by the constancy of the heat of activation. If a temperature independent viscosity η' is defined as

$$\eta' = \eta / e^{\Delta H_v / RT}, \tag{8}$$

the families of viscosity *vs* shear stress curves (Fig. 6) can be reduced to one curve for each resin. Furthermore, it has been suggested that at low shear stresses these curves approach a constant value (Newtonian behavior), and it is reasonable to assume that a Newtonian component of the observed viscosity is present at all shear stresses but can be detected only when its contribution is comparatively large, i.e., at low shear stresses. On this basis the following equation can be written, using temperature independent values,

$$1/\eta' = (1/\eta_0') + (1/\eta_s'), \tag{9}$$

where η' = observed viscosity; η_0' = Newtonian viscosity; η_s' = shear stress dependent viscosity. Thus, a viscosity *vs* shear stress curve can be divided into the two above components. Performing this operation on the temperature independent viscosity *vs* shear stress curve for resin *A* produces the two curves shown, along with the original curve, in Fig. 9. The curve for the shear stress dependent component evidently follows the equation,

$$\eta_s' = KS^{-n}, \tag{10}$$

where K, n = constants. A similar situation was found to apply for resins *B* and *C*. Combining Eqs. (9) and (10) yields

$$1/\eta' = (1/\eta_0') + (S^n/K). \tag{11}$$

This same type of equation has been found applicable by Rabinowitsch,[10] Bestul,[14] and Ferry.[15]

Equation (10) describes the function for the stress dependent viscosity, and this can be combined with the stress function for the elastic shear strain [Eq. (7)] to produce

$$\eta_s' = KS_0^{-n} e^{-(n/c)\gamma_e}. \tag{12}$$

Thus there exists a unique relation between the stress dependency of the viscosity and the magnitude of the elastic shear strain.

It will now be shown that the stress dependency of the viscosity, as described by Eq. (12), fits in well with the concept of an entropy effect. That such an effect is active might be suspected for two reasons. First, according to Mack,[16] the applicability of a power function [Eq. (10)] indicates the change in viscosity to be attributable to an entropy effect. Second, the shape of the shear stress *vs* elastic shear strain curve (Fig. 7) is similar to that for rubber-like elasticity which phenomenon, on the basis of kinetic theory, is explainable in terms of an entropy effect.[17]

It is common practice to express the variation in viscosity in terms of a free energy of activation[12] giving rise to the expression

$$\eta = C \exp(\Delta F^{\ddagger}/RT), \tag{13}$$

where F = Gibbs free energy and $\Delta F^{\ddagger}$ = free energy change associated with the activation of flow. Using the thermodynamic equivalence for the free energy change

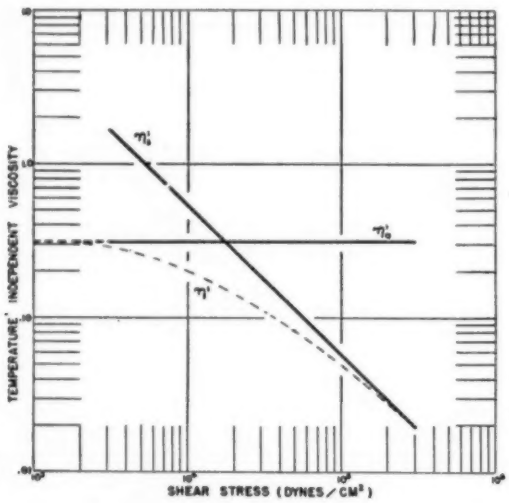


FIG. 9. η', η_0', and η_s' *vs* shear stress curves for polyethylene resin *A*.

[12] Glasstone, Laidler, and Eyring, *Theory of Rate Processes* (McGraw-Hill Book Company, Inc., New York, 1941).

[13] H. M. Smallwood, J. Appl. Phys. 8, 505 (1937).

† Data obtained by extrusion plastometer techniques may show an effect of the type discussed. However, it should be noted that such data are frequently the result of both viscous and elastic properties, and thus theories dealing with viscosity alone are inapplicable.

[14] Bestul, Belcher, Quinn, and Bryant, J. Phys. Chem. **56**, 432 (1952).

[15] J. D. Ferry, J. Am. Chem. Soc. **64**, 1330 (1942).

[16] C. Mack, paper presented at the Society of Rheology Meeting, Chicago, Illinois, October 27, 1951.

[17] See for example L. G. R. Treloar *The Physics of Rubber Elasticity* (Clarendon Press, Oxford, 1949).

Eq. (16) can be written

$$\eta = C \exp(\Delta H^{\ddagger}/RT - \Delta S^{\ddagger}/R), \quad (14)$$

where $\Delta H^{\ddagger}$ = heat (or enthalpy) of activation and $\Delta S^{\ddagger}$ = entropy of activation. The implication of this is more clearly seen if Eq. (12) is restated as

$$\eta_s = K S_0^{-n} e^{\Delta H_v/RT - n\gamma_e/c}. \quad (15)$$

The similarity between Eqs. (14) and (15) suggests that the stress dependency of the viscosity is the result of an entropy change whose value is $n\gamma_e R/c$.

V. ACKNOWLEDGMENTS

The helpful comments of the writer's associates, particularly those of Dr. D. I. Marshall, are gratefully acknowledged.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 9 SEPTEMBER, 1954

Internal Friction and Dynamic Modulus of Cold-Worked Metals*

A. S. Nowick
Hammond Metallurgical Laboratory, Yale University, New Haven, Connecticut
(Received April 12, 1954)

The complexities of the internal friction and dynamic modulus of metals in the cold-worked condition may be understood reasonably well by separation of the phenomena into the following three distinct effects: (1) the "nonlinear" effect, characterized by strong amplitude dependence, frequency independence, and its disappearance after severe deformation; (2) the Köster effect, which shows rapid recovery after deformation, at temperatures well below the recrystallization temperature; (3) the "viscosity" effect, characterized by strong frequency and temperature dependence. The origin of each of these effects is briefly discussed from the point of view of dislocation theory.

INTRODUCTION

THERE has been much confusion in the interpretation of internal friction phenomena in cold-worked metals due to the difficulty in establishing the basic experimental facts. For example, conflicting observations are reported with regard to the dependence of the internal friction and dynamic modulus on the frequency and strain amplitude of oscillation, and on the temperature. There are also varied observations on the dependence of the damping on the degree of deformation; several observers report a maximum in this functional relationship while in other cases only a monotonic increase is obtained. Studies of the effects of annealing at "low" temperatures (below the recrystallization temperature of the metal) upon the internal friction after cold working provide further conflicting data. The present paper will attempt to show that most of the confusion in the study of these damping phenomena can be resolved by subdividing the behavior of deformed metals into three distinct effects. The special characteristics of each effect which distinguish it from the others are first discussed. The final "Discussion" section deals briefly with the mechanisms of the various effects in terms of dislocation theory.

Simultaneously with the measurement of internal friction, one may obtain the dynamic modulus M from the resonant frequency of vibration of a given specimen. (Young's modulus is obtained for longitudinal or flexural vibration, while the dynamic shear modulus is measured when torsional oscillations are used.) A change in internal friction is usually accompanied by a corresponding change in the dynamic modulus; the portion of the dynamic modulus which is closely related to the internal friction may be called the ΔM *effect*.[1] In order to define this quantity we consider that the strain ϵ at any instant during the vibration cycle when the stress is σ is composed of two parts,

$$\epsilon = \epsilon' + \epsilon''. \quad (1)$$

Here,

$$\epsilon' = \sigma/M' \quad (2)$$

is the perfectly elastic strain which obeys Hooke's law, while ϵ'', the nonelastic strain, results from internal rearrangement and is not simply proportional to the stress. Equation (2) may be regarded as the definition of M'. It is readily shown[1] that if ϵ_1' is the amplitude of the elastic strain and ϵ_1'', ϵ_2'' the amplitudes of the components of the nonelastic strain in phase with and out of phase with the stress, respectively, then the internal friction is given by

$$\tan\phi = \epsilon_2''/\epsilon_1'. \quad (3)$$

The quantity ϕ, the angle by which the total strain lags behind the stress, is a convenient measure of the internal friction. Usually ϕ is sufficiently small that $\tan\phi \cong \phi$. The dynamic modulus, on the other hand, is generally lower than M' by an amount

$$\Delta M = M' - M, \quad (4)$$

* Work supported by the Office of Ordnance Research, U. S. Army.

[1] A. S. Nowick, *Progress in Metal Physics IV*, B. Chalmers, editor (Interscience Publishers, Inc., New York, 1953), Chap. I.

given by

$$\Delta M/M' = \epsilon_1''/\epsilon_1'. \quad (5)$$

This lowering of the elastic modulus due to the presence of nonelastic strain is the ΔM effect.

There are two general types of mechanisms which give rise to nonelastic strain. In the first, or *viscoelastic* case, ϵ'' is time dependent as a result of relaxation and flow phenomena. The quantity ΔM is then a function of frequency such that $\Delta M \to 0$ as $\omega \to \infty$ (ω = angular frequency). The quantity M' may therefore be defined in operational terms as the limiting value of the dynamic modulus, M, as $\omega \to \infty$. In the second or *static hysteresis* case, ϵ'' is independent of the frequency but not single valued. Thus the hysteresis loop traversed during one oscillation of stress is independent of the rate of traversal. The principal difference between the two cases is that in the first, hysteresis occurs because of the *dynamic* nature of the experiment and the time dependence of ϵ'', while in the second, the hysteresis loop is characteristic of the behavior of the material, in that it appears equally under semistatic and dynamic conditions. The static hysteresis loop is characteristically nonlinear, and therefore the internal friction and dynamic modulus are usually strongly amplitude dependent,[2] even at very low strain amplitudes where viscoelastic effects would normally be independent of amplitude. For the static hysteresis case, the ΔM effect is still defined by Eq. (5), but in order that this definition be unambiguous it is most convenient to take $\Delta M = 0$ at the limit of zero amplitude of vibration.

In general the behavior of a material may involve complex combinations of viscoelasticity and static hysteresis. The quantity M' may then be defined as the dynamic modulus in the limit of zero amplitude and infinite frequency; ΔM is then defined by Eq. (4).

From an atomistic viewpoint we may not wish to call M' the true modulus of the material if this quantity contains terms other than those resulting from the mean interatomic forces. For the case of crystals, such a situation arises when dislocations or segments of dislocations, capable of oscillating completely in phase with the stress,[3] are present. Such oscillation of dislocations can contribute to the total elastic strain ϵ' according to

$$\epsilon' = \epsilon_i + \epsilon_d, \quad (\epsilon_d \ll \epsilon_i), \quad (6)$$

where ϵ_i is a measure of interatomic forces and ϵ_d is due to dislocation oscillations. From Eq. (6) it follows that M', or more conveniently the compliance M'^{-1}, may be regarded as composed of two parts,

$$M'^{-1} = M_i^{-1} + M_d^{-1}, \quad (7)$$

in which M_i is the *true modulus* and M_d the contribution due to the oscillation of dislocations.

[2] Equations (3) and (5) show that the quantities ϕ and $\Delta M/M'$ are independent of the amplitude of strain if ϵ_2'' and ϵ_1'' are proportional to ϵ_1', i.e., if the equation relating nonelastic strain and stress is linear.

[3] J. Friedel, Phil. Mag. **44**, 444 (1953).

PROPERTIES OF ANNEALED METALS

As a basis for reference in the discussion of internal friction and dynamic modulus in cold-worked metals, the nature of these same properties in well annealed metals will first be reviewed. Polycrystalline metals show the well-known grain boundary peak[4] in the plot of internal friction as a function of temperature, which is believed to arise as a result of the relaxation of shear stress across grain boundaries. The large breadth of this internal friction peak suggests a wide distribution of relaxation times. This result is understandable in terms of the distribution in dimensions of the grain boundaries. Other effects in polycrystalline annealed metals include a rise in internal friction[5,6] at high temperatures (above the grain boundary peak) which has not been studied extensively, and a thermoelastic effect[7] due to the flow of heat currents between grains. The latter is usually insignificant unless the grain size is close to a critical value.

In single crystals, low frequency measurements (i.e., with a torsion pendulum) do not detect significant effects up to high temperatures, presumably because of the relatively large background damping in the apparatus. With high frequency methods, and the lower background attainable, a strongly amplitude-dependent internal friction and ΔM effect is found in the vicinity of room temperature for a wide variety of crystalline materials. Since this effect is increased in magnitude by small amounts of deformation, it is discussed in the following section, as the first of three effects which contribute to the internal friction of cold-worked metals.

THE "NONLINEAR" EFFECT

Various investigators[8] of damping phenomena at frequencies in the tens of kilocycles have shown that for both well annealed and slightly cold-worked metal single crystals, the internal friction is a strongly increasing function of the amplitude of vibration, even at strain amplitudes as low as 10^{-7}. It is very common for such measurements to show that the internal friction at strain amplitude of 10^{-6} is increased over the value at zero amplitude by a factor of two, and in some cases by nearly one hundred. Correspondingly, the dynamic modulus decreases with increasing strain amplitude. (Although single crystals are used in most of the investigations of this effect, the same behavior is observable in polycrystalline metals as well.) Such an amplitude dependent internal friction and ΔM effect implies[2] that the relation between stress and non-

[4] T. S. Kê, Phys. Rev. **71**, 533 (1947); **72**, 41 (1947).

[5] T. S. Kê, J. Appl. Phys. **21**, 414 (1950).

[6] Pearson, Greenough, and Smith, Nature **167**, 1023 (1951).

[7] C. Zener, *Elasticity and Anelasticity of Metals* (The University of Chicago Press, Chicago, 1948).

[8] Papers are given in reference 1.

elastic strain is nonlinear even at very small strains. As noted in the Introduction, such nonlinear behavior would be expected to characterize static hysteresis, rather than a viscoelastic mechanism. This belief is supported by the fact that in all known cases in which relaxation-type phenomena occur in metals, internal friction is practically independent of amplitude at least up to strain amplitudes of 10^{-5}. A quantitative study[9] of the ratio of the amplitude dependent part of the internal friction to $\Delta M/M$, as a function of frequency and temperature, confirms the static hysteresis mechanism. This mechanism is further supported by the fact that, according to the most reliable information available, the amplitude dependent damping is independent of frequency.[9,10]

This "nonlinear" effect is very sensitive to a small amount of cold working, even that which may be produced by handling. The working produced by the application of static stress which is insufficient to cause large-scale yielding results, nevertheless, in a very large increase in the internal friction of single crystals of copper.[9,11] In the case of brass single crystals, which show a sharply defined yield point, no significant increase in internal friction is produced through the application of stresses below the yield point.[12] It therefore appears that some deformation, even though not readily detectable by crude measurements, is required in order that an increase in the nonlinear damping effect be produced.

The application of a large static stress well above the yield point of a metal seems to destroy the "nonlinear" effect, as demonstrated by Weertman and Koehler[13] for single crystals of copper. After such deformation the damping is observed to be almost insensitive to amplitude up to strain amplitudes greater than 10^{-6}.

The variation with temperature of the amplitude-dependent internal friction and dynamic modulus has been studied to a limited extent.[9,14] These nonlinear phenomena show only a mild temperature dependence by comparison to diffusion controlled viscoelastic phenomena, but lack of agreement on the detailed temperature dependence indicates that further work is necessary.

Inasmuch as small amounts of cold working greatly increase the amplitude dependent internal friction, it appears significant to inquire whether a tendency exists for the damping to recover in time toward its original value. Although this information is not reported explicitly in the literature dealing with the "nonlinear" effect, it appears from the author's own results and from the work of others that for copper, the major part of the increased damping produced by deformation does not recover at room temperature over several days. (The situation is different for zinc,[11] which is able to recrystallize at room temperature.)

TABLE I. Characteristics of the three effects responsible for the internal friction of cold worked metals.

	Nonlinear	Köster	Viscosity
Amplitude dependence at low amplitude	Strong	None or slight	None
Degree of prior deformation required	None or slight	Moderate to large	Large
Frequency dependence	None	Mild	Strong
Temperature dependence	Mild	Mild	Strong
Recovery below recrystallization temperature	Slight	Complete	Limited

The principal characteristics of the "nonlinear" effect, as well as those of the two remaining effects to be discussed in succeeding sections, are summarized in Table I.

THE KÖSTER EFFECT

Experiments by Köster and collaborators[15,16] reveal an internal friction effect which seems to be of a different nature from that discussed in the previous section. The studies have been made on polycrystalline brass and iron in the kilocycle range of frequencies.[17] It is found that after moderate or severe deformation high values of internal friction are observed at room temperature immediately after the deformation. This internal friction shows a strong tendency to recover *completely* at temperatures well below that required for recrystallization.[18] Figure 1 illustrates this recovery phenomenon at room temperature; the data show almost a complete disappearance of the effect in deformed iron in about ten hours. For higher temperature anneals (e.g., 50 to 100°C) the effect disappears much faster. The large increase in damping immediately after the deformation is accompanied by a decrease in dynamic modulus of several percent. This change in modulus breaks into two parts. The first part recovers in about the same time as the internal friction; the second is a residual effect which shows no recovery, even for anneals at elevated temperatures, until recrystallization occurs. The partial recovery of the modulus is illustrated in Fig. 1. The residual effect is apparent through the difference between the final values for the 80 percent and 25 percent deformed metal. This internal friction phenomenon and the recoverable part of the change in

[9] A. S. Nowick, Phys. Rev. **80**, 249 (1950).
[10] E. P. T. Tyndall, unpublished results on zinc crystals. The frequency dependence of this internal friction effect is difficult to measure because of the extreme structure sensitivity of the property (see reference 9).
[11] T. A. Read, Trans. Am. Inst. Mining Met. Engrs. **143**, 30 (1941).
[12] G. H. Found, Trans. Am. Inst. Mining Met. Engrs. **161**, 120 (1945).
[13] J. Weertman and J. S. Koehler, J. Appl. Phys. **24**, 624 (1953).
[14] J. Pittinger, Phys. Rev. **83**, 872 (1951).
[15] W. Köster and K. Rosenthal, Z. Metallkunde **30**, 345 (1938).
[16] W. Köster, Arch. Eisenhüttenw. **14**, 271 (1940–41).
[17] Similar experiments on aluminum are complicated by still another effect and will therefore not be discussed until the next section.
[18] Thus the recovery of internal friction takes place without any detectable change in the hardness, tensile strength, or microstructure of the metal.

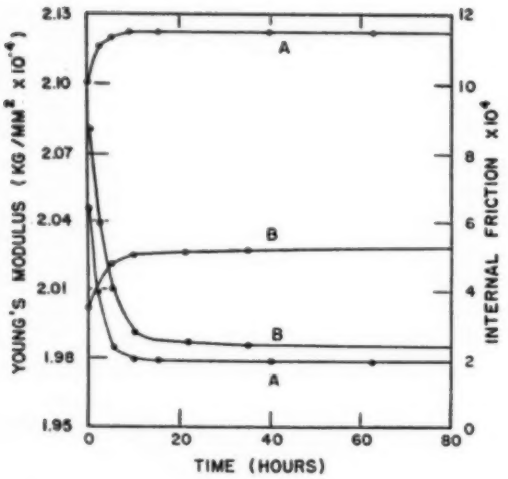


FIG. 1. Internal friction (solid circles) and dynamic Young's modulus (open circles) of Armco iron, measured at 20°C as a function of time after deformation. Degree of cold drawing: *A*—25 percent, *B*—80 percent. (After Köster.)

modulus constitute what we shall call the "Köster effect."

The present effect differs from the "nonlinear" effect not only in the larger degree of deformation required and in the rapid recovery, but also in that it is not strongly amplitude dependent. The evidence for this latter statement is not entirely direct, since the magnetic method used by Köster could probably only attain strain amplitudes in the range 10^{-7}. However, in another paper[19] it appears that amplitude dependence could readily be observed with this equipment for the case of ferromagnetic metals (which show strong nonlinear behavior as a consequence of the existence of ferromagnetic domains[1]). More direct evidence is obtained from low frequency torsion pendulum measurements on cold-worked metals at strain amplitudes of about 10^{-5}, carried out by Boulanger[20] (on a Ni-Cr alloy) and in this laboratory[21] (on platinum wire). These experiments, which presumably measure the same Köster effect as obtained at higher frequencies, also indicate that the effect is not strongly amplitude-dependent. This apparent linearity of the effect suggests that it is probably viscoelastic in origin. This statement cannot be made with certainty, however, since little is known about the dependence of the effect on frequency and temperature. It appears however that the Köster effect is not strongly dependent on frequency and temperature. This statement is based first, on the fact that it is observable with about the same magnitude over a wide frequency range and second, on the experiments of Hasiguti[22] who reports that internal friction of heavily cold-worked copper is practically independent of temperature in the range 0°C to −100°C. These results imply that if the Köster effect is caused by a viscoelastic phenomenon, a wide distribution of relaxation times is involved.

Köster reports that the initial internal friction increases monotonically with the degree of cold working. This result differs from that of Hasiguti and Hirai[23] who obtain a maximum in the dependence of damping on deformation. Unfortunately, these latter authors do not obtain the amplitude dependence of the internal friction which would make it possible to determine whether the "nonlinear" effect enters into their measurements. In view of the large degree of deformation at which some of the maxima are obtained and the fact that the internal friction is recoverable[23] below the recrystallization temperature, it seems that the internal friction reported by these authors is predominately the Köster effect. Part of the discrepancy between observers may be related to the very rapid recovery that takes place in the first few minutes after deformation. Thus, what is reported as "initial damping" depends on how much time is lost before measurements begin.

The magnitude of the recoverable part of the dynamic modulus is of the order of several percent, depending on the degree of deformation. In addition to the experimental work already quoted, there is the work of Smith[24] on polycrystalline copper, which deals with detailed measurements of the dynamic Young's modulus without reporting corresponding observations of internal friction. There are two possible interpretations of the recoverable part of the modulus. First, it might be supposed that since modulus and damping recover together, the increase in modulus is a ΔM effect, as defined in the Introduction. This point of view was previously adopted by the writer.[1] Second, a substantial part of the recoverable modulus change may be a dislocation contribution to M', as discussed in relation to Eq. (7). Inasmuch as measurements of this effect over a very wide range of frequencies are not available, an attempt may be made to distinguish between these two possibilities by examination of the ratio of the recoverable part of the internal friction to the recoverable frictional change in modulus. From the values reported by Köster[15,16] this ratio is consistently of the order of 0.01. It would not be possible to explain such a relatively large modulus change, i.e., such a small ratio, in terms of a ΔM effect, unless internal friction takes on very high values at low temperatures.[25] Since there seems to be no evidence for this latter possibility, it is more reasonable to believe that the second of the two interpretations given above is the correct one.

[19] F. Förster and W. Köster, Naturwiss. **25**, 436 (1937).
[20] C. Boulanger, Compt. rend. **226**, 1170 (1948).
[21] A. S. Nowick and A. E. Roswell (unpublished work).
[22] R. R. Hasiguti (private communication).
[23] R. R. Hasiguti and T. Hirai, J. Appl. Phys. **22**, 1084 (1951).
[24] A. D. N. Smith, Phil. Mag. **44**, 453 (1953).
[25] This conclusion is arrived at by means of a relation (see reference 7, page 55) which enables one to obtain the ΔM effect (for viscoelastic behavior) at a given temperature by integration of the internal friction curve from absolute zero to that temperature.

THE "VISCOSITY" EFFECT

When a metal is heavily deformed, a residual internal friction is observed after the Köster effect has completely recovered, provided that the measurements are made at sufficiently high temperature or low frequency. Thus, Köster finds[26] that deformed aluminum shows a different behavior from that of higher melting-point metals. He finds that room temperature internal friction still remains at a high level after recovery has taken place and that this residual effect cannot be removed completely until the specimen is annealed above the recrystallization temperature. It may then be expected that the residual effect is strongly temperature-dependent, and that it appears in aluminum, and not in brass or iron, only because room temperature is higher on an appropriate "reduced temperature" scale for the case of aluminum. This viewpoint is substantiated by the fact that measurements of internal friction as a function of temperature[15,26] show a sharply increasing curve for cold-worked aluminum and not for brass.

This residual effect increases rapidly with decreasing frequency and is therefore most conveniently studied by low frequency (torsion pendulum) methods. The most complete study of this effect is that of Kê[27] and Zener[28] on aluminum, in which low frequency and static measurements are utilized. Since a long time elapsed between the deformation and the taking of measurements, the Köster effect may be regarded as having recovered completely. Figure 2 shows typical curves illustrating the strong temperature dependence. Each curve is obtained after a one-hour anneal at the highest temperature of measurement. It is clear that partial recovery takes place at each successively higher annealing temperature. Recrystallization occurs at about 300°C after which the grain boundary peak, characteristic of a well-annealed polycrystalline metal, is obtained. The dependence of internal friction on frequency and temperature may be expressed in the usual way as a dependence on the parameter $\omega e^{H/RT}$, where H is the heat of activation. For the case of cold-worked aluminum $H=31$ kcal/mole. By utilization of static measurements at constant temperature and this value of the heat of activation, it is possible to obtain the extension of the internal friction curves of Fig. 2 to higher temperatures without the additional complication of recovery. For example, creep measurements show that the 250°C curve would attain a value as high as $\phi=0.6$ at 425°C if it could be extended to this temperature without the occurrence of a change in condition (recovery) of the metal due to the higher temperature. Such a continuously rising internal friction curve is suggestive of a viscous material, hence the name "viscosity effect."

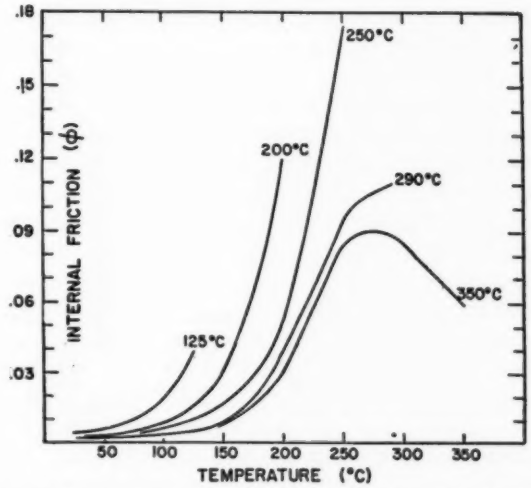


FIG. 2. Dependence of internal friction of heavily cold worked aluminum on the temperature of measurement, after a series of anneals. The corresponding annealing temperature is marked at the top of each curve. (After Kê.)

The work of Kê[27] shows that for the internal friction effect under discussion, there is no amplitude dependence up to strain amplitudes of 10^{-5}. A ΔM effect undoubtedly accompanies the internal friction, but no systematic measurements of the dynamic modulus are reported.

DISCUSSION

The preceding sections separate the internal friction of cold-worked metals into three categories. This classification is based entirely on phenomenological grounds, as summarized in Table I. The present section will briefly describe what seem to be the most reasonable atomistic interpretations of the three effects.

The "nonlinear" effect has been described in terms of a static hysteresis mechanism. The experimental evidence indicates that the sources which can be taken through a static hysteresis loop are present in well-annealed single crystals and may be increased in numbers by minute amounts of deformation; however, they are rendered ineffective once severe deformation takes place. It seems natural to consider that the dislocation segments which generate slip at high stress levels, the so-called Frank-Read generators,[29] are the sources capable of undergoing hysteresis at low stress levels. These dislocation segments are capable of bowing out under an applied stress and will oscillate if the stress is periodic. In order to account for internal friction it is necessary to assume that each source is restrained by barriers (e.g., solute atoms) which prevent it from following the stress completely in phase. For each increment of stress, there is a corresponding number of dislocations which are torn loose and move rapidly to a point where the motion is again interrupted

[26] W. Köster, Z. Metallkunde **32**, 282 (1940).
[27] T. S. Kê, Trans. Am. Inst. Mining Met. Engrs. **188**, 575 (1950).
[28] T. S. Kê and C. Zener, Symposium on the Plastic Deformation of Crystalline Solids (U. S. Office of Naval Research, Pittsburgh, 1950), p. 185.
[29] W. T. Read, Jr., *Dislocations in Crystals* (McGraw-Hill Book Company, Inc., New York, 1953).

by an appropriate barrier. This motion of a dislocation is almost instantaneous, each "burst" to a new position being accompanied by a small increment in the nonelastic strain. These displacements are not completely reversed upon removal of the stress i.e., stress of opposite sign is required before dislocations are restored to their original positions, because of the restraining effect of the barriers; therefore a stress-strain hysteresis loop is obtained. If a crystal does not show a well defined yield point (as is the case in copper), the application of static stress below the macroscopic yield point will result in some generation of new dislocation loops. These loops may participate in the hysteresis phenomenon, so that the effect increases. When large scale plastic flow occurs, however, the interaction of the generated loops upon each other and upon the original source produce a large "force constant" for dislocation movement; a small oscillatory stress will therefore no longer produce hysteresis.

The effect of severe deformation is to produce fragmentation of the lattice into "crystallites"[30] or "subgrains."[31] The formation of this substructure is believed to be responsible for the increased hardness and strength of cold-worked metals. This substructure is not strongly affected by low temperature annealing but is removed only by recrystallization. The rapid recovery of the Köster effect well below the recrystallization range suggests that this effect is produced by dislocations found within the subgrains. Such dislocations are relatively free (i.e., not closely interlaced); they are therefore capable not only of producing internal friction but also of contributing to M' through purely elastic oscillation under the influence of the applied stress [Eq. (7)]. The rapid recovery shows that these dislocations are unstable and suggests that they tend to move to positions where they are more tightly bound.[32]

The large residual lowering of the elastic modulus, which does not recover below the recrystallization temperature (see Fig. 1), is apparently a lowering of the true modulus M_t due to the decrease in the average interatomic force constants as a result of the deformation. This interpretation is in accord with the conclusion of Zener[33] that internal strains, i.e., the existence of atoms displaced from normal lattice positions, always result in a lowering of the average tensile or shear moduli.

Finally, the "viscosity" effect may be related to the subgrain structure since it is only completely obliterated by recrystallization. According to Zener[7] this effect is due to viscous relaxation of shear stress across internal surfaces formed by the deformation, in a manner analogous to grain boundary relaxation. The principal difference between the two effects is that the grain boundary relaxation produces an internal friction peak, while the "viscosity" effect does not. This difference may be attributed[1,28] to the irregularity of the network of subgrain boundaries, in contrast to the more regular network formed by the grain boundaries. In terms of dislocation theory the relaxation centers created by plastic flow are the dense collections of dislocations in the subgrain boundaries. A detailed atomistic interpretation of the mechanism of relaxation does not yet exist. However, based on the results for aluminum, it appears that the heat of activation for both the "viscosity" effect and for grain boundary relaxation are, within experimental error, the same as the heat of activation for lattice self diffusion. It therefore seems that the rate-controlling process of both effects involves a diffusive flow of atoms through the lattice.

Because the "viscosity" effect is interpreted in terms of dense collections of dislocations, the detailed dislocation mechanism of this effect may not be as readily obtainable as for the other two effects discussed here. Conversely, further study of the "nonlinear" and Köster effects will probably yield the most useful information about the behavior of dislocations, since these effects depend more nearly on the movement of individual dislocations.

[30] W. A. Wood, Proc. Roy. Soc. (London) **A172**, 231 (1939).
[31] A. Guinier, *Imperfections in Nearly Perfect Crystals*, W. Shockley, editor (John Wiley and Sons, Inc., New York, 1952), Chapter 15.
[32] A more complete analysis of these recovery phenomena will be presented in a forthcoming paper.
[33] C. Zener, Acta Cryst. **2**, 163 (1949).

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 9 SEPTEMBER, 1954

The Rheology of Suspensions

KEITH H. SWEENY AND RICHARD D. GECKLER
Aerojet-General Corporation, Azusa, California
(Received June 7, 1954)

The rheological properties of concentrated suspensions of sized glass spheres in an equal-density medium has been investigated with the rotational viscosimeter. It has been found that the variation of fluidity with rate of shear can be expressed satisfactorily by a form of an expression suggested by Reiner relating the fluidity to the shearing stress, the fluidity at rest, and the fluidity at infinite shearing stress. It is also shown that values of the apparent fluidity are influenced by the time of shearing, the aging of the suspensions, and temperature effect attributed to the work of shearing.

The viscosity of suspensions of sized spheres at concentrations as great as 55 volume percent solids may be satisfactorily represented by either the relation of Mooney or of Maron, Madow, and Krieger; the identity of their relations is also shown. The constants for the equation relating viscosity and concentration are shown to be in satisfactory agreement with those predicted by Mooney. Constants for the viscosity-concentration equation are also given for several examples of experimental work selected from the literature; the form of equation proposed by Mooney relating viscosity and concentration is generally useful though the constants may not agree with the predicted values.

It has been found that the apparent fluidity of the glass bead suspensions decreases with an increase in particle diameter when the suspending fluid is zinc bromide in aqueous glycerol though the viscosity is independent of particle diameter when the suspending fluid is the nonaqueous mixture of ethylene tetrabromide and diethylene glycol.

The viscosity of a bimodal system of glass beads was shown to decrease very significantly as the bulk density was increased by packing progressively smaller spheres into the interstices of the larger spheres of the bimodal system, in qualitative agreement with a relation proposed by Mooney.

I. INTRODUCTION

THE rheological properties of suspensions are becoming of increasing interest in connection with many industrial problems. However, despite the attention given this subject by scientific workers in the past, few experimental data exist for simple systems defined by a small number of parameters. Available data on numerous different suspensions exhibit such a variety of departures from Newtonian behavior that it is difficult to characterize their behavior in any simple way. For this reason the authors undertook a study of suspensions containing uniform glass spheres in an equal-density medium of zinc bromide dissolved in aqueous glycerol. The relation between the shearing stress and rate of shear, the effect of concentration on fluidity, the variation of fluidity with particle size, and the effect of bimodal distribution of particle sizes on fluidity were determined from measurements using a rotational viscosimeter. Although the rheological behavior turned out to be more complicated than originally anticipated, the data clearly illustrate many of the fundamental types of rheological properties.

II. EXPERIMENTAL METHODS AND MATERIALS

In order that the relationship between fluidity and rate of shear over an extended range of rate of shear values might be adequately studied, a rotational viscosimeter of the Couette type was chosen for the investigation. In this apparatus the sample is contained between two coaxial cylinders. The outer cylinder is capable of being rotated so as to give any of a large range of values for the rate of shear at the inner cylinder. The shearing stress can be expressed as a function of the restoring moment required for nonrotation of the inner cylinder.

The rotational viscosimeter* used in this investigation was similar to that described by Green.[1] A view of the instrument showing the driving unit, the torsion spring for measuring shearing stress, and the sample cup and bob is shown in Fig. 1. The unit was powered by a $\frac{1}{4}$ horse power synchronous motor; the speed of rotation of the cup was controlled by a Graham vari-speed drive. The speed of rotation of the cup was indicated by the gauge attached to the speed control (near the center of Fig. 1). The speed control was carefully calibrated and is checked at reasonable intervals. It is estimated that the reproducibility of resetting the speed control is 0.1 rpm or better for the range of 0 to 440 rpm. The constancy of speed with rotation is excellent: no change in rpm was noted in a period of over two hours continuous rotation. The instrument was supplied with six torsion springs and three sizes of cups and bobs. The three sizes of cups and bobs were supplied so that an attempt could be made to evaluate a possible slippage layer at the walls of the cup and bob.

The shearing stress, P, in dyne cm^{-2} was calculated from the relation

$$P=\frac{M\theta}{2\pi R_i^2 h}, \tag{1}$$

where M is the spring constant (dyne cm deg^{-1}), θ is the deflection of the spring in degrees, R_i is the radius

* Constructed by Instrument Development and Manufacturing Company, Pasadena, California.

[1] H. Green, *Industrial Rheology and Rheological Structures* (John Wiley and Sons, Inc., New York, 1949), pp. 100–111.

FIG. 1. Aerojet rotational viscosimeter.

of the bob in cm, and h is the depth of immersion of the bob, corrected for traction at the end of the bob by the method of Lindsey and Fischer.[2] It has been found convenient to let

$$K=\frac{M}{2\pi R_i^2 h}, \tag{2}$$

so that

$$P=K_2\theta. \tag{3}$$

A series of six springs with spring constants ranging from 179.0 to 34 550 dynes deg^{-1} were used. The shearing stress constant, K_2, varied from 3.882 to 749.3 dynes cm^{-2} deg^{-1} for the cup and bob combination used in Procedures 1 and 2 (see below). The value of K_2 ranged from 2.041 to 394.0 for the cup and bob used in Procedure 3.

The rate of shear, V, in sec^{-1}, was given by the relation

$$V=\frac{4\pi n}{1-a}, \tag{4}$$

where n is the revolutions per second and a is an instrumental constant equal to $(R_i/R_0)^2$, where R_i is the radius of the bob and R_0 is the radius of the cup. Letting

$$K_1=\frac{4\pi}{1-a}, \tag{5}$$

we have

$$V=K_1 n. \tag{6}$$

The value of K_1 was held at 0.840 sec^{-1} min for all work except that designated as Procedure 3 (see following). The rate of shear constant K_1 was 1.098 sec^{-1} min for the cup and bob used in Procedure 3.

[2] C. H. Lindsey and E. K. Fischer, J. Appl. Phys. 18, 988 (1947).

The glass spheres used in the preparation of the suspensions were closely sieved fractions of commercial fine glass beads.† Microscopic examination of the sieved samples revealed that the range of particle diameters was within 5 to 10 percent of the average particle diameter and were at least 90 percent spheres. The particle-size distributions of the specimens are given in Table I.

A suspending medium with density equal to that of the glass spheres was chosen in order to preclude sedimentation. Following the work of Vand,[3] a solution of commercial fused zinc bromide in aqueous glycerol was chosen as the suspending medium. Stock solutions used in this investigation generally were prepared with the following composition:

	Wt percent
Zinc bromide	76.1
Glycerol, U.S.P.	5.3
Water	18.6
	100.0

This solution had a density of 2.494 g ml^{-1} and behaves as a simple Newtonian fluid with a viscosity of 2.70

TABLE I. Particle-size distribution of glass spheres used in rheological studies.

Glass spheres separated by sedimentation

D_{vs}[a] 12.6 microns		D_{vs} 35.9 microns		D_{vs} 124 microns	
Distribution microns	%	Distribution microns	%	Distribution microns	%
0–4	21.0	0–10	1.0	88–100	3.0
5–8	28.0	11–20	8.0	101–108	11.5
9–12	31.5	21–30	22.0	109–116	11.0
13–16	15.0	31–40	53.5	117–124	23.5
17–20	4.5	41–50	15.5	125–132	20.5
				133–140	23.5
	100.0		100.0	141–148	6.0
				>149	1.0
					100.0

Glass spheres separated by sieving

D_{vs} microns[b]	Retained sieve mesh	microns	Passed sieve mesh	microns
49	325	44	270	53
97	170	88	150	105
116	150	105	120	125
164	100	147	80	177
226	65	210	60	239[c]
262	60	250	48	282[c]

Density of all fractions of glass spheres except the 12.6 micron sample was 2.600 g ml^{-1}. The 12.6 micron fraction had a density of 2.746.

[a] D_{vs} = volume-surface mean diameter.
[b] The volume-surface mean diameter of the sieved beads was calculated by assuming an even distribution of sizes in the range separated by the sieves. All sieve fractions were examined microscopically to verify the sieve limits.
[c] The diameters in microns are those of the manufacturer's maximum listed size for these fractions. Tests at Aerojet showed that all beads passed the sieve size listed. The mesh sizes were 250 and 297 microns.

† Vendors: Potter Brothers, Ozone Park, New York, and Minnesota Mining and Manufacturing Company, Scotchlite Division, St. Paul, Minnesota.

[3] V. Vand, J. Phys. & Colloid Chem. 52, 300 (1948).

poise at 25°C. The solution is sufficiently nonhygroscopic so that stock solutions may be prepared and stored provided that the solution is tightly stoppered and handled with care. The solution viscosity can be varied by changing the glycerol-water ratio, should that be desired.

It was desired for some experiments that a nonaqueous suspending medium be used. This medium was prepared with the following composition:

	Wt percent
Ethylene tetrabromide	91.8
Diethylene glycol	8.2
	100.00

This solution had a density of 2.599 g ml^{-1} and behaves as a simple Newtonian fluid with a viscosity of 0.157 poise at 25°C.

The experimental determinations were generally made under one of the following three schedules of determinations. The schedules have been established from a practical consideration of the various factors which might influence the consistency data.

Procedure 1

Rate of shear 0 to 147 sec^{-1}, in 16 steps of increasing rate followed by 16 steps of decreasing rate, using cup of inside radius 1.5988 cm and 5.099 cm depth. The bob had a radius of 1.2987 cm and was immersed to a depth of 3.987 cm (4.216 cm, corrected for end effect). Samples used for this determination were aged approximately 20 min.

Procedure 2

Rate of shear 0 to 366 sec^{-1}, in 12 steps of increasing rate followed by 12 steps of decreasing rate using same cup and bob dimensions as described above. Samples to be determined according to Procedure 2 were aged approximately 40 min and were remixed after 20–30 min aging.

Procedure 3

Rate of shear 0 to 192 sec^{-1}, in 16 steps of increasing rate of shear followed by 16 steps of decreasing rate using cup of inside radius 1.9987 cm and depth 5.100 cm. The bob had a radius of 1.7980 cm and was immersed to a depth of 3.988 cm (4.369 cm, corrected for end effect). Samples used for Procedure 3 were aged approximately 60 min and were remixed after 20–30 min and 40–50 min aging. Readings during all determinations were made at 30-sec intervals. The consistency determinations were made with separate portions of the same sample and were run in the order listed. Samples run by Procedure 1 were aged the least period of time and are believed to be the most reliable. All determinations were made at 25.0±0.1°C.

III. RELATION BETWEEN SHEARING STRESS AND RATE OF SHEAR

The purpose of studies with the rotational viscosimeter was to determine the complete relation between the shearing stress p (dyne cm^{-2}) and the rate of shearing strain $\dot{e}$ (sec^{-1}) for the material in question; the rotational viscosimeter is particularly useful in obtaining this variation of viscosity with changes involving a wide range of rate of shear values. For a simple Newtonian fluid, the relation is

$$\dot{e}=\varphi p, \tag{7}$$

where φ is the fluidity (cm^2 $dyne^{-1}$ sec^{-1} or $poise^{-1}$), and is independent of $\dot{e}$ and p. For concentrated suspensions, however, φ is not a constant and it must be expressed as a function of either $\dot{e}$ or p. Since neither the shearing stress p nor the rate of shearing strain $\dot{e}$ can be generally measured, it is desirable to define an apparent fluidity φ' in terms of the consistency variables V and P.

$$V=\varphi' P, \tag{8}$$

where V is the rate of shear (sec^{-1}) at the wall of the inner cylinder and P is the shearing stress (dyne cm^{-2}). The shearing stress, P, is proportional to the restoring moment on the inner cylinder and the rate of shear, V, is proportional to the angular velocity of the outer cylinder. The proportionality constants can be calculated from the instrumental dimensions and fluidity relations can be calculated without recourse to artificial standards.

For suspensions of solid particles, four different phenomena can complicate the simple behavior summarized in Eq. (7) or (8). These are (a) dependence of fluidity on shearing stress, (b) dependence of fluidity on duration of shearing stress and age of samples, (c) temperature effects of shearing, and (d) slippage at the walls of the coaxial cylinders.

The dependence of fluidity on shearing stress means that φ in Eq. (7) must be a function of p. This function must account for the fact that φ is a minimum (φ_0) at zero rate of shear and rises to a maximum (φ_∞) as the rate of shear is increased. Reiner[4] has proposed two functions that satisfy these requirements,

$$\varphi=\frac{\varphi_0+\varphi_\infty p^2/\alpha^2}{1+p^2/\alpha^2}, \tag{9}$$

and

$$\varphi=\varphi_\infty-(\varphi_\infty-\varphi_0)\exp-p^2/\beta^2. \tag{10}$$

In Eq. (9), α is the shearing stress at which φ is halfway between φ_0 and φ_∞, and in Eq. (10), β (β is $\sqrt{\chi}$ in Reiner's treatment) represents the shearing stress at which φ is 63.2 percent of the way from φ_0 to φ_∞. When Eq. (9) is expressed in terms of the consistency vari-

[4] M. Reiner, *Deformation and Flow* (H. K. Lewis and Company, Ltd., London, 1949), pp. 108–120.

ables, V and P, this becomes

$$V=\varphi_\infty P-\frac{\alpha(\varphi_\infty-\varphi_0)}{1-a}\tan^{-1}\frac{1-a}{aP/\alpha+\alpha/P}. \quad (11)$$

Substituting the apparent fluidity φ' for V/P, then Eq. (11) may be expressed as

$$\varphi'=\varphi_\infty-\frac{\varphi_\infty-\varphi_0}{(1-a)P/\alpha}\tan^{-1}\frac{1-a}{aP/\alpha+\alpha/P}. \quad (12)$$

Expanding in terms of $\tan^{-1}x$, dropping items after the first, and rearranging, Eq. (12) can be replaced with less than 1 percent error by

$$\varphi'=\frac{\varphi_0+\varphi_\infty aP^2/\alpha^2}{1+aP^2/\alpha^2}, \quad (13)$$

which is a functional expression for the apparent fluidity of the same form as Eq. (9) with $P\sqrt{a}$ replacing p.

In terms of the consistency parameters Eq. (10) becomes

$$V=\varphi_\infty P-\frac{\beta\sqrt{\pi}(\varphi_\infty'-\varphi_0)}{2(1-a)}\left(\operatorname{erf}\frac{P}{\beta}-\operatorname{erf}\frac{aP}{\beta}\right), \quad (14)$$

where $\operatorname{erf}x$ is the tabulated value of the Gauss error function

$$\operatorname{erf}x=\frac{2}{\sqrt{\pi}}\int_0^x \exp-x^2dx. \quad (15)$$

The data obtained in the experimental determinations were used to evaluate the parameters in the equations above. The initial fluidity was obtained by plotting the apparent fluidity, φ', as a function of the square of the shearing stress and extrapolating to zero shearing stress. The apparent fluidity at infinite shearing stress was obtained from an extrapolation of the plot of apparent fluidity as related to the reciprocal of the shearing stress.

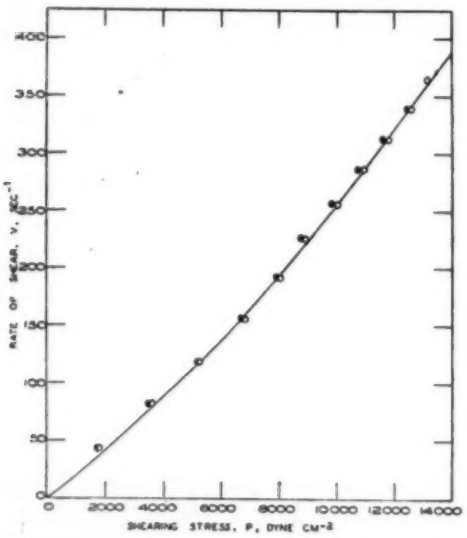


FIG. 2. Shearing stress vs rate of shear for 45 volume percent suspension of 49-micron glass spheres in zinc bromide, water, glycerol mixture. Experimental data for increasing rates of shear are represented by open circles, while the data obtained at decreasing rates of shear are represented by closed circles. The line was calculated from Eq. (13).

Equations (13) and (14) have been used to fit experimental data in a number of instances. It has been found that Eq. (13), an integrated form of Eq. (9), presents the data slightly more accurately than the more cumbersome Eq. (14), an integrated form of Eq. (10). The data shown by Fig. 2 indicate the accuracy with which an experimental consistency curve for increasing rates of shear can be fit by Eq. (13).

The experimental data depart from the calculated consistency curve at very low rates of shear as can be seen from the first two points of Fig. 2 (the divergence at maximum shearing stress is because of experimental error). The divergence of the values at low rates of shear is always toward a higher apparent fluidity. Although these results might be explained by the lower precision of the shearing stress measurements at low rates of shear where the torsion spring deflections are small, this explanation is not believed satisfactory. In addition, a wall slippage-layer at low rates of shear would not explain the results. The discrepancy between experimental and calculated values at very low rates of shear is being investigated to see if the behavior is being complicated by some previously unrecognized phenomenon.

It is concluded that the variation of apparent fluidity with increasing rates of shear may be adequately represented by Eq. (9), except for deviations at low rates of shear. This enables one to mathematically relate the variation of apparent fluidity with experimentally determined values of the initial fluidity, the fluidity at infinite shearing stress, and the shear stress corresponding to a mean fluidity of the initial and final fluidities.

The factor of time affects the value of the fluidity in two ways. First, the sample when sheared would exhibit thixotropic breakdown to some extent. Examples of this are given in Tables II and III. Another phenomenon which was noted was the factor which is termed "aging." Although the stress decay is a reversible phenomenon, the factor of "aging" is nonreversible and cannot be changed, for example, by remixing the suspensions. The possible sedimentation of the particles can be shown by the following calculation. The sedimentation of spheres during a determination can be calculated from Stokes' Law, modified for concentration effects by the method of Steinour.[5] The 49-micron diameter glass beads suspended in the zinc bromide-aqueous glycerol medium of density 2.612 g ml^{-1} and viscosity 1.606 poise (Table II) would rise but 7 microns in a 20-minute period during the test. This represents an effective concentration change of but 0.006 percent in the annular space between the cup and bob. It is therefore believed that

[5] H. H. Steinour, Ind. Eng. Chem. 36, 618 (1944).

TABLE II. Effect of aging on the shearing stress at constant rate of shear.

40 vol percent suspensions of 49-micron glass spheres in zinc bromide, glycerol, and water mixture[a]

Time at constant rate of shear[b] sec	Shearing stress, dynes cm^{-2} de-aerated samples aged, 10 min	1 hr[a]	3 hr[c]
10	2644	2575	2133
30	2610	2555	2137
60	2610	2548	2178
120	2616	2555	2260
180	2623	2568	2308
600	2678	2596	2473

50 vol percent suspensions of 49-micron glass spheres in ethylene tetrabromide-diethylene glycol[d]

Time at constant rate of shear[e] sec	Shearing stress, dynes cm^{-2} de-aerated samples aged, 20 min	2 hr[a]
10	448	423
30	444	423
60	427	415
120	421	405
180	411	401
600	387	390

[a] Density of medium 2.612 g ml^{-1}, viscosity 1.606 poise.
[b] Rate of shear = 147.1 sec^{-1}.
[c] The samples were remixed after aging in order to assure elimination of concentration gradients in the suspensions.
[d] Density of medium 2.599 g ml^{-1}, viscosity 0.157 poise.
[e] Rate of shear 100.8 sec^{-1}.

the phenomena of stress decay and aging cannot be associated with settling. The phenomenon of aging is not understood but is believed associated with adsorption of the suspending fluid on the spheres. It can only be concluded that the time of shearing and the aging of the suspensions must be rigorously controlled.

It has been found that the temperature effects on shearing are very important in establishing the relation between shearing stress and rate of shear. Since heat is generated in viscometric measurements, a decrease in viscosity will occur upon shearing which is dependent upon the temperature coefficient of viscosity of the suspension, the apparatus dimensions, the rate of shear, the viscosity of the suspensions, and the thermal properties of the cup containing the sample and bath. All of the results reported in this paper have been corrected for the effects of temperature by the method of Weltmann and Kuhns.[6] It has been assumed in these calculations that an infinite heat transfer coefficient exists at the outer surface of the cup. The magnitude of the corrections can be seen from the following summary:

Viscosity, poise	Rate of shear, sec^{-1}	Δt, °C	$\Delta\eta$, poise
10	147	0.032	0.03
10	366	0.189	0.19
20	147	0.064	0.13
20	366	0.378	0.76
50	147	0.160	0.80
50	366	0.945	4.72

If there should be slippage at the walls of the coaxial cylinders, the rheological behavior of the samples would be the same as though a thin layer of pure suspending fluid covered the walls. The apparent fluidity would then be a function of the instrumental dimensions and probably also of the rate of shear.

Table III shows a comparison of results for a suspension of glass spheres measured in three different sizes of cups and bobs. These cups and bobs were selected such that the instrumental constant (R_i^2/R_0^2), where R_i is radius of the bob and R_0 is radius of the cup; the a value of Eq. 3 *et seq.* remained the same for the three sizes of cups and bobs. It is concluded that differences in Table III are due to experimental errors and that no wall slippage layer was evident in this investigation. Vand[3] found a wall-slippage layer whose thickness is 1.08 times the radius of the particles. The application of his value to the results of this investigation yields values for viscosities which are unrealistically large.

Although the variables noted in the preceding paragraphs affect consistency measurements, they can be effectively controlled and in some cases measured. The temperature effects of shearing can be calculated. By making the consistency determinations with a controlled time cycle of measurements and to a specified maximum rate of shear, it is possible to measure the extent of thixotropic breakdown, since the extent of thixotropic breakdown will be quantitatively represented, for a thixotropic substance, by the area enclosed by the curves for increasing and decreasing rates of shear. The curves for increasing and decreasing rates of shear of a thixotropic substance are said to describe the thixotropic hysteresis loop. Uniform aging of the suspensions has been used in making determinations of consistency.

IV. RESULTS

The results of the determinations of fluidity are given in Table IV; the particle-size analysis of the spheres is given in Table I. The determination of the rheological parameters has been made at concentrations of from

TABLE III. Effect of duration of stress and instrument dimensions on the shearing stress at constant rate of shear for 40 vol percent suspensions of 124-micron glass spheres.

(All samples de-aerated and aged 1 hr[a] before determination)

Rate of shear sec^{-1}	Time after start, sec	Shearing stress, dynes cm^{-2} bob 1[b]	bob 2[c]	bob 3[d]
53.8	30	637	674	657
53.8	180	619	659	657
53.8	600	614	655	647
100.8	30	1315	1403	1276
100.8	180	1274	1384	1283
100.8	600	1247	1376	1248
147.1	30	1644	1733	1690
147.1	180	1712	1733	1701
147.1	600	1760	1729	1680

Suspending medium: zinc bromide-glycerol-water mixture density 2.578 g ml^{-1}, viscosity 1.411 poise.

[a] Samples were remixed after being aged 50 min in order to preclude sedimentation.
[b] Bob 1: radius 1.2987 cm.
[c] Bob 2: radius 1.7318 cm.
[d] Bob 3: radius 1.7980 cm.

[6] R. N. Weltmann and P. W. Kuhns, J. Colloid Sci. 7, 218 (1952).

TABLE IV. Values of apparent fluidity at rest, apparent fluidity at infinite shearing stress, and empirical parameters α and β for suspensions of glass beads in zinc bromide, glycerol, and water solution.

(Fluidity of medium =0.3701 poise^{-1}, density 2.494 g ml^{-1})

Concentration vol %	Range, rate of shear, sec^{-1}	D_m, microns	Apparent fluidity at rest φ_0 poise^{-1}	Apparent fluidity at inf. shear. stress φ_∞, poise^{-1}	α dynes cm^{-2}	β dynes cm^{-2}
20.00	0–147	49	0.1773	0.1852	1275	1400
20.00	0–366	49	0.1840	0.1947	1890	2500
20.00	0–192	49	0.1848	0.1965	1345	1970
20.00	0–147	116	0.1772	0.2043	1320	1520
20.00	0–366	116	0.1900	0.1977	2320	3120
20.00	0–192	116	0.2010	0.2270	1080	1510
32.00	0–147	49	0.0670	0.0761	2820	4000
32.00	0–366	49	0.0732	0.0753	4440	5880
32.00	0–192	49	0.0888	0.0925	3900	5000
32.00	0–147	116	0.0745	0.0806	2270	2560
32.00	0–366	116	0.0836	0.0849	3600	4450
32.00	0–192	116	0.0741	0.0790	3120	4340
40.00	0–147	36	0.0393	0.0464	6020	6580
40.00	0–366	36	0.0446	0.0515	8780	12 350
40.00	0–192	36	0.0423	0.0488	4670	6330
40.00	0–147	49	0.0398	0.0500	5880	7690
40.00	0–366	49	0.0438	0.0534	10 000	13 500
40.00	0–192	49	0.0400	0.0508	6250	8330
40.00	0–147	116	0.0448	0.0536	3330	4540
40.00	0–366	116	0.0512	0.0588	7300	10 000
40.00	0–192	116	0.0428	0.0553	4170	5720
40.00	0–147	226	0.0566	0.0573	4200	5210
40.00	0–366	226	0.0538	0.0639	7700	10 310
40.00	0–192	226	0.0537	0.0600	4760	7700
44.83	0–147	49	0.0196	0.0256	7000	7850
44.83	0–366	49	0.0204	0.0340	11 300	12 640
44.83	0–192	49	0.0231	0.0309	7700	10 530
45.00	0–147	124	0.0253	0.0361	5490	7350
45.00	0–366	124	0.0338	0.0455	9430	12 980
45.00	0–192	124	0.0323	0.0470	5620	8000
45.00	0–17.2	124	0.0350	0.0382	682	760
50.00	0–147	49	0.00940	0.0170	8860	9900
50.00	0–192	49	0.01615	0.0234	10 630	14 920
50.00	0–17.2	49	0.00790	0.0136	2000	2320
50.00	0–147	124	0.01366	0.0277	8190	11 240
50.00	0–366	124	0.02415	0.0366	13 980	18 860
50.00	0–192	124	0.01790	0.0318	9090	12 120
55.00	0–17.2	49	0.00595	0.00683	2750	3400
55.00	0–22.5	49	0.00652	0.00774	3300	4100
55.00	0–147	124	0.00870	0.0205	12 350	17 230
55.00	0–192	124	0.00880	0.0234	14 080	18 860
55.00	0–17.2	124	0.00584	0.00757	2450	2850
55.00	0–22.5	124	0.00723	0.00782	3700	4900
60.00	0–17.2	124	0.00344	0.00346	4900	4900

20 vol percent to as high a solid content as could be measured with monodisperse systems. When the solid content of the suspensions is increased further, insufficient liquid is present to form a smooth suspension and falsely high values of apparent fluidity are obtained. Measurements have been made with two sizes of spheres over the continuous range of concentrations, and at 40 vol percent, the determinations were made with four fractions of glass spheres in order that particle size effects might be evaluated.

A. The Variation of Viscosity with Concentration of the Suspension

The first attempt at expressing viscosity as a function of concentration is the familiar Einstein equation

$$\eta_r = 1 + kC, \tag{16}$$

where η_r is the relative viscosity and C is the volume concentration. The factor k depends upon the shape of the particle and is equal to 2.5 for spheres. This relation is applicable only to very dilute suspensions, however. Numerous attempts have been made to extend this equation to higher concentrations of suspended solids, generally by the addition of higher order empirical constants.

One of the most significant advances in the expression of viscosity at higher concentrations is due to Vand,[7] who derived Eq. (11), which holds at much higher concentrations than the results of previous workers, from theoretical considerations,

$$\ln\eta_r = \frac{k_1C + r_2(k_2-k_1)C + \cdots}{1-QC}. \tag{17}$$

In this equation $k_1=2.5$ is the Einstein shape factor of single spheres, $k_2=3.175$ is the shape factor of collision doublets, $r_2=4$ is the collision time constant, and $Q=0.609$‡ is the hydrodynamic interaction constant. In the present investigation, the equation due to Vand was found to express adequately the variation of viscosity with concentration for values as high as 30 volume percent solids. However, the equation developed by Vand has disadvantages, it does not take into account particle size, particle-size distributions, nor interparticle forces which give rise to non-Newtonian behavior or thixotropy.

There have been several other relations presented which are useful for suspensions as concentrated as 20 to 30 volume percent solids. These include the relations of Roscoe,[8] who has presented a relation which considers size distribution as a variable, Oliver and Ward,[9] and Simha.[10] The relation presented by Simha was developed from hydrodynamic considerations. However, from the considerations of simplicity and usefulness over a range of concentration up to 50 volume percent or more, the most satisfactory relation for expressing the variation of viscosity with concentration has been that due to Mooney[11] and to Madow and Krieger.[12,13] The general expressions for polydisperse suspensions proposed by Mooney is:

$$\ln\eta_r = 2.5\sum_{i=1}^{n}\frac{C_i}{1-\sum_{j=1}^{n}\lambda_{ij}C_j}, \tag{18}$$

where η_r is the relative viscosity of the suspension, C_i is the volume fraction of the ith particle size, and there

[7] V. Vand J. Phys. and Colloid Chem. **52**, 277 (1948).

‡ M. Mooney (see reference 11) believes that the value of Q should be somewhat larger.

[8] R. Roscoe, Brit. J. Appl. Phys. **3**, 267 (1952).

[9] D. R. Oliver and S. G. Ward, Nature **171**, 396 (1953).

[10] R. Simha, J. Appl. Phys. **23**, 1020 (1952).

[11] M. Mooney, J. Colloid Sci. **6**, 162 (1951).

[12] I. M. Krieger and S. H. Maron, J. Colloid Sci. **6**, 528 (1951)

[13] S. H. Maron, B. P. Madow and I. M. Krieger, J. Colloid Sci. **6**, 584 (1951).

are n different particle sizes. The terms subtracted from 1 in the denominator explain self-crowding and crowding among particles of different sizes and hydrodynamic interaction. This relation can take into account the variation in particle size and in particle-size distribution and the crowding constants can be modified by the inter-particle forces responsible for non-Newtonian behavior or for thixotropy.

The relation due to Mooney becomes

$$\ln\eta_r = 2.5C/(1-kC), \tag{19}$$

where k is the self-crowding factor and hydrodynamic interaction factor, for the special case of the viscosity of a monodisperse suspension. A convenient expression for Eq. (15) is to rearrange in the following form:

$$1/\ln\eta_r = -1/\ln\varphi_r = b(-k+1/C). \tag{20}$$

A plot of $1/\ln\eta_r$ *vs* $1/C$ will be a straight line of slope b and intercept bk. Mooney gives the value 1/2.5 or 0.400 for b, based on the Einstein shape factor for dilute suspensions of spheres.

The relation due to Maron, Madow and Krieger[12,13] is given by

$$\log\eta'_r = bZ, \tag{21}$$

where η'_r is the apparent relative viscosity, b is a constant and Z is a concentration factor represented by the relation

$$Z = \alpha C/(1-\alpha C), \tag{22}$$

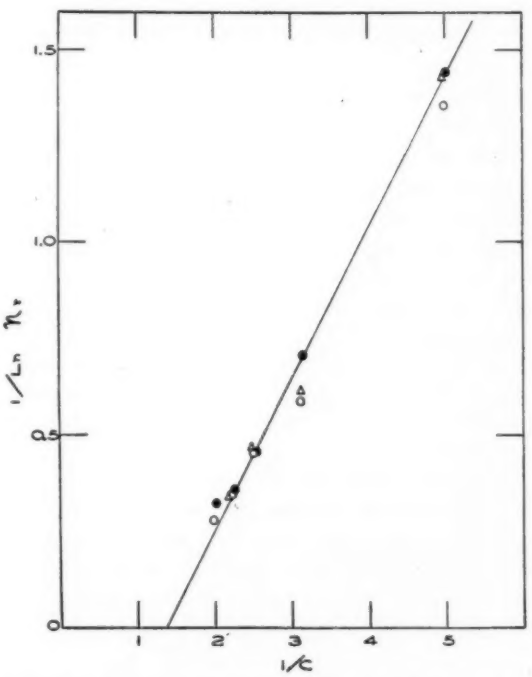


FIG. 3. Reciprocal of natural logarithm of relative viscosity *vs* reciprocal of concentration as volume fraction for suspensions of 49-micron glass spheres in zinc bromide, glycerol, water mixtures. Relative viscosity calculated from values of apparent fluidity at rest. Rate of shear: ○ Procedure 1. △ Procedure 2. ● Procedure 3.

TABLE V. Values of constants k and b for Eq. (20) relating viscosity and concentration for sized glass spheres in a zinc bromide, water and glycerol medium.

Particle size D_{st}, microns	Range, rate shear, sec^{-1}	Range, concentration, vol %	η_r calculated from φ_0		η_r calculated from φ_∞	
			k	$2.5b$	k	$2.5b$
49	0–147	20–50	1.356	0.977	1.364	1.007
	0–366	20–45	1.470	1.010	1.244	1.033
	0–192	20–50	1.338	0.980	1.113	0.990
116–124[a]	0–147	20–55	1.267	0.946	1.073	0.935
	0–366	20–50	1.204	0.993	1.083	1.016
	0–192	20–55	1.268	1.030	1.000	1.030

[a] Two separate fractions of glass spheres were used in making this series of determinations, one having a volume-surface mean diameter of 116 microns, the other 124 microns. These determinations have been combined to give the data listed as 116–124 microns particle diameter.

where α is a constant and C is the volume fraction of solids. Maron, *et al.* prefer to rearrange Eq. (21) after substituting Eq. (22) to give the following form:

$$C/\log\eta'_r = 1/\alpha b - C/b. \tag{23}$$

Thus a graph of $C/\log\eta'_r$ *vs* C should be linear with slope $-1/b$ and intercept $1/\alpha b$, so that both α and b can be evaluated. It can be shown that the relation of Eq. (23) can be rearranged to the form of Eq. (20), for which the equivalent constants are as follows:

Equation	(20)	(23)
Slope	1/2.5	$1/2.303\alpha b$
Intercept	$-k/2.5$	$-1/2.303b$

The results of a test of Eq. (20) are shown in Fig. 3. It was found that linear relations were obtained for the plot of $1/\ln\eta_r$ *vs* $1/C$ in all of the series of determinations; Eq. (20) can be fitted using the values of the relative viscosity at rest or at infinite shearing stress. A summary of the values of the constants for Eq. (20) when applied to the data of this investigation is given in Table V. It will be seen from the tabulated values of $2.5b$ that the slope of the curves varies from 93.5 to 103.3 percent of the theoretical value of 0.400 given by Mooney. The agreement is considered to be a satisfactorily good check of the theoretical value. Mooney believes that the self-crowding factor should be between the limits 1.35 and 1.91; the self-crowding factor in this work varies from 1.000 to 1.470, values somewhat less than would be expected by theory.

In order to test the limits for the constants suggested by Mooney, data have been collected from several sources and the constants have been obtained by plotting the data in the form of Eq. (20). The results are given by Table VI.

The suspensions prepared by Eilers[14] were composed of fine bituminous particles (2.7 and 2.2 microns average diameter) suspended in water. The viscosities of the suspensions were measured in a special type of capillary pipette capable of giving the viscosity at several rates of shear. The constants for Eilers' results have both a

[14] H. Eilers, Kolloid-Z. **97**, 313 (1941).

TABLE VI. Comparison of constants for Eq. (20) calculated from the data of various investigators.

Investigator	Type of suspension	Slope	Intercept
Theory by Mooney		0.400	−0.540 to −0.764
Eilers	Bituminous emulsion *A*	0.448	−0.400
	Bituminous emulsion *B*	0.471	−0.461
Maron, *et al.*	Latex suspensions, high shear stress	0.455	−0.641
	Latex suspensions, low shear stress	0.403	−0.609
Robinson	Glass spheres in motor oil	0.375	−0.400
	Glass spheres in motor oil	0.347	−0.350
	Glass spheres in castor oil	0.350	−0.310
	Glass spheres in glucose solution	0.287	−0.210
	Glass spheres in polyethylene glycol	0.221	−0.031
	Glass spheres in sucrose solution	0.268	−0.190
This work	Glass spheres 49μ, 0–147 sec^{-1}, φ_0	0.391	−0.530
	49μ, 0–366 sec^{-1}, φ_0	0.404	−0.593
	49μ, 0–192 sec^{-1}, φ_0	0.392	−0.525
	124μ, 0–147 sec^{-1}, φ_0	0.378	−0.480
	124μ, 0–366 sec^{-1}, φ_0	0.397	−0.478
	124μ, 0–192 sec^{-1}, φ_0	0.412	−0.522
Vand	Glass spheres in ZnI_2, glycerol, water	0.400	−0.584

greater slope and intercept than the theoretical Mooney values.

The data of Maron, Madow and Krieger[13] are given for a series of suspensions of latex particles diluted with water. Maron *et al.* found, it necessary to use two flow equations to fit the data for solid contents above 35 vol percent. Two sets of constants are therefore given for the equations, one set for values of the shearing stress lower than a critical value at concentrations above 35 vol percent solids, and for the entire shearing stress region at lower values of the solid concentration. The second set of values of the constants represent the data at concentrations from 35 to 62 vol percent solids with values of the shearing stress in excess of the critical value.

The determinations of Robinson[15] were made using a rotational viscosimeter. The glass spheres used in the suspensions ranged from 10 to 30 microns in diameter, though mainly between 10 and 20 microns. The glass spheres were suspended in a variety of media, as shown by Table VI. Although the values of the slope with non-polar suspending media approach the theoretical values of Mooney, the constants for the flow data when the spheres were suspended in sugar solutions or in polyethylene glycol are significantly lower in slope than the value predicted by Mooney. The data given by Robinson show some divergence from simple Newtonian behavior at high rates of shear; an average viscosity was used in his work, rather than expressing viscosity as a function of shearing stress.

The results of Vand[3] were for suspensions of glass spheres having a diameter of 100 to 160 microns. Measurements at concentrations from 35 to 54 percent by volume were made with a rotational viscosimeter while other measurements as high as 50 vol percent were made using modified Ostwald viscosimeters. The data of Vand were corrected for a wall slippage layer 1.08 times the average particle radius in thickness. The description of Vand's experiments indicates behavior which could be attributed to thixotropy at concentrations greater than 40 vol percent. The constants for Eq. (20) are obtained from smoothed values of the data from both rotational and capillary viscosimeters; the slope and intercept compare very well with the values predicted by Mooney.

The results of the tabulated experiments show that the form of Eq. (20) is generally applicable to viscosity data over a wide range of concentrations. It is apparent, however, that the predicted values for the constants are not generally followed. There are several considerations which might account for this. The slope has been derived from the low concentration Einstein factor for spheres. Departure from spherical shape for the particles, departure of the consistency curves from simple Newtonian behavior, and the temperature effect on shearing would all be expected to influence the value of the slope. It is known that the particles used in this investigation were spherical in shape and the flow curves could not be expressed by simple Newtonian behavior, though the use of viscosity parameters at the standard conditions of shear (at rest and infinite shearing stress) seems to have overcome this difficulty. The intercept of Eq. (20) will also be influenced by the departure from simple Newtonian flow. The self-crowding and inter-crowding effects will be influenced by the packing of the solid particles. The packing, in turn, will be related to the particle-size distribution of the particles of solid, and variation from the predicted values of Mooney will be expected.

It must be concluded that the form of Eq. (20) is satisfactory for expressing the variation of viscosity with concentration but that the constants may depart from the predicted values because of departure from simple Newtonian behavior, of inter-particle forces, and of variation in particle size distribution.

B. Effect of Particle Diameter on Rheological Properties

The theoretical equations evolved for the viscosity of suspensions, from the early equation of Einstein to

[15] J. V. Robinson, J. Phys. & Colloid Chem. 53, 1042 (1949).

the recent results of Vand,[7] have not considered particle size as a variable. However, it has been observed that the apparent fluidity for a given concentration of glass spheres suspended in the zinc bromide, glycerol, and water medium decreased as the diameter of the beads was decreased. Figure 4 shows the effect of particle size on the apparent fluidity at infinite shearing stress for 40 vol percent suspensions of sized glass spheres in the zinc bromide-aqueous glycerol medium. The particle-size analysis for the spheres is given in Table I and the designation of procedures is the same as that used in the discussion of the effect of concentration on viscosity. Curves similar to those of Fig. 4 are obtained for the apparent fluidity at rest *vs* the logarithm of the volume-surface mean diameter of the particles. The values of the empirical parameters α and β (Eqs. 9–14) used in describing variation of apparent fluidity with shearing stress apparently do not depend on particle size, though they do depend upon concentration.

When the zinc bromide, glycerol, and water medium is replaced by the nonaqueous ethylene tetrabromide-diethylene glycol medium, however, the effect of particle size on the apparent fluidity at rest and the apparent fluidity at infinite shearing stress vanishes. Results are shown in Fig. 4 of the apparent relative fluidity at infinite shearing stress for 50 vol percent suspensions of the sized glass spheres in the nonaqueous medium. The apparent relative fluidity was used instead of the apparent fluidity with this curve as a convenient transformation of coordinates to allow all the data to be shown on one graph.

There are several possible phenomena which might account for the observed effect of particle size on flow properties when the suspending medium is changed.

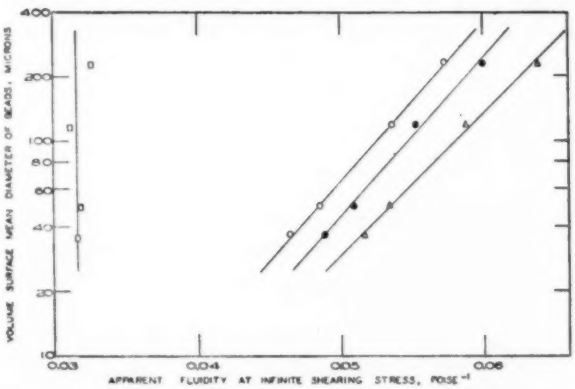


FIG. 4. Apparent fluidity infinite shearing stress *vs* volume surface mean diameter of glass spheres. Samples designated Procedures 1, 2, and 3 were prepared as 40 volume percent suspensions in zinc bromide, glycerol water mixture of density 2.494 g ml^{-1}, fluidity 0.3701 poise^{-1}. The sample designated Procedure 4 was prepared as a 50 volume percent suspension in the ethylene tetrabromide-diethylene glycol mixture of density 2.599 g ml^{-1}, fluidity 6.37 poise^{-1} and the determinations were made with the same rate of shear cycle (0–147 sec^{-1}) as Procedure 1. The apparent relative fluidity is shown with the Procedure 4 sample. Rate of shear: ○ Procedure 1. △ Procedure 2. ● Procedure 3. □ Procedure 4.

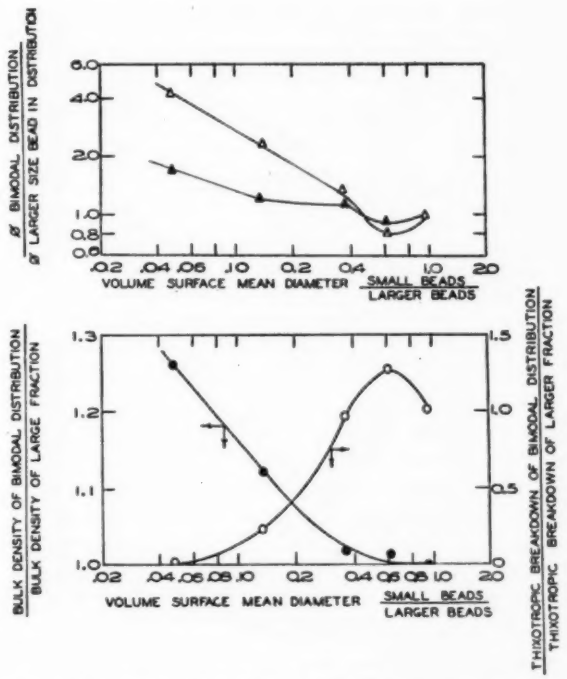

VOLUME SURFACE MEAN DIAMETER SMALL BEADS / LARGER BEADS
BULK DENSITY OF BIMODAL DISTRIBUTION / BULK DENSITY OF LARGE FRACTION
THIXOTROPIC BREAKDOWN OF BIMODAL DISTRIBUTION / THIXOTROPIC BREAKDOWN OF LARGER FRACTION

FIG. 5. Relative fluidity of bimodal distributions of 55 volume percent suspensions of glass spheres in zinc bromide, water, glycerol mixture. Rate of shear 0–147 sec^{-1}. Composition of samples given in Table VII. △ Apparent fluidity at rest. ▲ Apparent fluidity at infinite shearing stress. ○ Ratio thixotropic breakdown. ● Ratio bulk density.

The observed effect that the apparent fluidity decreases as the particle-size decreases when the glass spheres are suspended in zinc bromide dissolved in aqueous glycerol might be explained by a wall-slippage layer. However, it was shown in Table III that the results of this investigation are apparently not affected by a slippage layer at the walls. Further evidence may be obtained from the results of Weltmann and Green,[16] who also observed that the fluidity decreased as the particle diameter decreased. The particles used by Weltmann and Green were so small (0.5 to 2.5 microns diameter) compared to the annular gap between the cup and bob that a wall-slippage correction calculated by the method of Vand[3] would be so small that it would be negligible. A further explanation of the observed decrease in fluidity of suspensions of glass spheres as the particle diameter of the spheres decreases is that to be attributed to an electroviscous effect. A double-ionic layer of the suspending medium around the nonconducting charged spheres in an electrolyte would increase the effective concentration of the particles, the increase being greatest for the smallest particles, and a result such as Fig. 4 might be obtained. The results of Weltmann and Green again preclude this possibility because a nonelectrolytic oil was used as the supporting medium for their investigation. A further possibility is that each glass sphere is surrounded by an adsorbed layer of fluid

[16] R. N. Weltmann and H. Green, J. Appl. Phys. **14**, 569 (1943).

TABLE VII. Fluidity of bimodal distributions of 55 vol percent suspensions of glass spheres[a] in a water solution of zinc bromide and glycerol.[b]

Diam of small spheres, microns	Vol-surf mean diam small spheres, microns	Ratio diam small to large spheres	Apparent fluidity at rest, φ_0, poise^{-1}	Apparent fluidity inf. shear. stress, φ_∞, poise^{-1}
239–282	261.6	1.000	0.00958	0.0182
147–177	164.0	0.627	0.00770	0.0168
88–105	97.0	0.371	0.0129	0.0213
20–45	35.9	0.137	0.0229	0.0223
<20	12.6	0.048	0.0415	0.0310

[a] Glass spheres consisted of 75 percent of 239–282-micron diam and 25 percent of diam shown in first column.
[b] Suspending medium: density 2.494 g ml^{-1}, fluidity 0.3701 poise^{-1}.

which increases its effective size. If this occurs, the resulting increase in the volume concentration of particles will be largest for the small particles and a result such as presented in Fig. 4 would be obtained. It is conceivable that a nonaqueous medium such as the ethylene tetrabromide-diethylene glycol mixture would adsorb differently than the aqueous zinc bromide-glycerol suspending medium. The variation of fluidity with particle diameter observed when the zinc bromide-aqueous glycerol medium is used is being investigated further in order that it may be clearly established why the apparent fluidity varies with particle diameter in this system.

C. Effect of Change in Particle-Size Distribution on Rheological Properties of Suspensions

Experience has shown that the viscosity of a suspension having a given percentage of solids may be decreased by using selected blends of particle sizes in place of a single particle-size fraction. Roscoe[8] has given a discussion of this effect. In addition, the theoretical developments presented either by Mooney[11] or by Maron, Madow, and Krieger[13] predict the decrease in viscosity with increased bulk density of the solids. Experimental data to show this effect are given in Fig. 5; the composition of the samples is given in Table VII, and the particle-size distribution of the glass spheres is given in Table I.

The composition of the samples was chosen to follow some of the considerations given by Hudson.[17] In a close-packed aggregate there are two types of interstices; the so-called "square hole" or a reversed spherical cube bounded by six spherical surfaces, and a "triangular hole" or a reversed spherical tetrahedron bounded by four spherical surfaces. One square and two triangular holes are associated with each sphere in close packing. In order to pack a single sphere into a square hole, the diameter of the smaller sphere must be less than 0.414 that of the larger. The diameter of the smaller spheres must be less than 0.225 of the larger in order to pack into a triangular hole. When the diameter of the smaller spheres is less than 0.155 that of the larger, the smaller spheres may continuously thread through the labyrinth. The diameters of the spheres have therefore been chosen so that the small sphere is successively too large to pack into a square hole, small enough to fill a square hole, small enough to fill a triangular hole or a square hole and barely able to thread through the pack, and the smallest size could easily thread through the labyrinth.

It will be seen from the results that, although the initial viscosity and extent of thixotropic breakdown first increase slightly when the sphere diameter ratio is 0.627, they then decrease steadily when diameter ratios are lower. The initial viscosity of the bi-disperse system with a diameter ratio of 0.048 has been decreased 4.3 times over that of the monodisperse system. The bulk density also increases with the decrease in viscosity. The fluidity at infinite shearing stress increases with the decrease in diameter ratio. The fluidity at infinite shearing stress remains more nearly constant with variation in the diameter ratio of the particles of the bimodal suspensions, though it also increases somewhat as the diameter ratio decreases.

Additional data are being collected to test the generalized Mooney relation, Eq. (18). It is to be concluded that the data presented give quantitative evidence of the effect of bimodal particle distributions on the viscosity of suspensions and qualitative proof of the interpretation of the crowding factors of Eq. (19).

ACKNOWLEDGMENT

The authors wish to acknowledge their indebtedness to Dr. A. O. Dekker for many helpful discussions during the progress of the investigation.

[17] D. Hudson, J. Appl. Phys. **20**, 154 (1949).

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 9 SEPTEMBER, 1954

Regular Contributed Original Research

Analysis of Systems Involving Difference-Differential Equations*

T. M. BURFORD
Department of Electrical Engineering, University of Wisconsin, Madison, Wisconsin
(Received September 11, 1953)

It is known that many physical systems, both time and space dependent, are described by difference-differential equations. Although the exact solutions to such equations are known to be unique, their determination usually involves a segmented type of solution that is very laborious to obtain. An approximate method of determining any number of roots is described. This approximate procedure allows forward gain and delay to be freely chosen and gives immediately the maximum stable value of forward gain. In many cases, especially where an optimum value of damping constant or forward gain is to be found, it appears that this approximate procedure is more easily applied than any previously described method.

I. INTRODUCTION

IN control systems work, where time is the independent variable, a dead-time delay is said to exist whenever an effect is initiated by a cause at a prior instant of time. For instance if pressure is transmitted hydraulically the output pressure will, ideally, be the same as the input pressure variation, but will be delayed by the time of transmission through the hydraulic line. A lossless transmission line of infinite length gives this same type of response and similarly the lumped parameter delay line may be made to give a close approximation to ideal delay. All of these cases may be described by a transfer function of the form $\epsilon^{-\tau s}$, where τ is the delay time. From the definition of the natural logarithm,

$$\epsilon=\lim_{n\to\infty}\left(1+\frac{s\tau}{n}\right)^{n/s\tau},$$

$$\epsilon^{-s\tau}=\lim_{n\to\infty}\left(1+\frac{s\tau}{n}\right)^{-n}=\lim_{n\to\infty}\left[\frac{1}{1+\dfrac{s\tau}{n}}\right]^{n}.$$

Therefore, a large number of R–C sections, each of transfer function $1/1+(s\tau/n)$, will approximate this type of response if buffer amplifiers are inserted between sections to eliminate interaction.

If the transmission medium is pneumatic, a transfer function of the form $\epsilon^{-\sqrt{\tau s}}$ applies. In such case the output is not identical to the input. This form is also obtained in an infinite length transmission line having shunt conductance and series inductance equal to zero. Therefore a large number of simple lags, *not* separated by buffer amplifiers, would approximate this response. This transfer function of "distributed lag" $\epsilon^{-\sqrt{\tau s}}$ may also be approximated by pure delay and a small number of simple lags.[1] Therefore, only the $\epsilon^{-\tau s}$ case will be considered hereafter.

Although dead-time delays are the commonest example, restriction to time-dependent systems is unnecessary, since a system is easily visualized in which an effect begins to act only when a space variable exceeds a certain number. This number might be termed a "space delay." The preceding is not intended to imply that time and space functions are analyzed identically, for while a time function is not affected by events in "future time," a space function might be affected (as in a loaded beam) by "future space."

II. ANALYSIS

The general form of a linear difference-differential equation is

$$\sum_{\mu=0}^{M}\sum_{\nu=0}^{M}a_{\mu\nu}\frac{d^{\mu}y(t-\tau_{\nu})}{dt^{\mu}}=h(t), \tag{1}$$

where

$$\frac{d^{\mu}[y(t-\tau_{\nu})]}{dt^{\mu}}=0; \quad t<\tau_{\nu}. \tag{2}$$

Under very mild conditions, it has been shown[2,3] that (1) possesses a solution

$$y(t)=\sum_{k=0}^{\infty}P_k\epsilon^{s_k t}, \tag{3}$$

where P_k is a polynomial in t and s_k is a zero of the characteristic equation

$$Z(s)=\sum_{\mu}\sum_{\nu}a_{\mu\nu}s^{\mu}\epsilon^{-\tau_{\nu}s}. \tag{4}$$

It might be noted that if, in (1), the summation over μ is deleted (1) becomes a standard finite-difference equation. If we particularize (1) by considering only one $\tau_{\nu}=\tau_k$ nonzero and assume all $a_{\mu k}=0$ but $a_{kk}=\alpha$, we

* This paper is based on a thesis presented by the author in partial fulfillment of the requirements for the M.S. degree in Electrical Engineering at the University of Wisconsin.

[1] R. C. Oldenbourg and R. Sartorius, *The Dynamics of Automatic Controls* (American Society of Mechanical Engineers, 1948), p. 88.
[2] E. M. Wright, Proc. Roy. Soc. Edinburgh **A62**, 387–393 (1949).
[3] E. F. Pitt, Proc. Cambridge Phil. Soc. **40**, 199–211 (1944).

have, using $p=d/dt$,

$$F(p)y(t)+\alpha p^k y(t-\tau)=h(t), \tag{5}$$

where $p^k y(t)=0$; $-\tau<t\leqq 0$. Then substituting the Taylor series development of $y(t-\tau)$,

$$F(p)y(t)+\alpha p^k\left[y-\tau py+\frac{\tau^2 p^2}{2}y-\cdots\right]=h(t), \tag{6}$$

we arrive at an operational representation of (5),

$$F(p)y(t)+\alpha\epsilon^{-\tau p}p^k y(t)=h(t). \tag{7}$$

A solution of (5), or even (1), may always be had by recalling from (2) that a delayed term of negative argument vanishes, therefore for $0<t<\tau$, (5) is a "normal" differential equation without delayed terms and may be solved for $y(t)$ valid for $0<t<\tau$. Now for $\tau<t<2\tau$, $p^k y(t-\tau)$ is determinate and may be inserted in (5) allowing a solution to be found for $\tau<t<2\tau$, and so on.

Although exact and straightforward, this segmented type of solution is very laborious to obtain. A considerable simplification[4] may be had by considering the transform of (5):

$$[F(s)+\alpha s^k\epsilon^{-\tau s}]Y(s)=H(s)+\psi(s) \tag{8}$$

where $\psi(s)$ contains the initial conditions.

Therefore,

$$\begin{aligned}Y(s)&=\frac{H(s)+\psi(s)}{F(s)+\alpha s^k\epsilon^{-\tau s}}\\&=\frac{H(s)+\psi(s)}{F(s)}\,\frac{1}{1+\dfrac{\alpha s^k}{F(s)}\epsilon^{-\tau s}}\\&=\frac{H(s)+\psi(s)}{F(s)}\sum_{i=0}^{\infty}\left(\frac{-\alpha s^k\epsilon^{-\tau s}}{F(s)}\right)^i.\end{aligned} \tag{9}$$

It is seen that the above series in powers of $\epsilon^{-\tau s}$ corresponds to a segmented time function as found before.

If in (7), $\epsilon^{-\tau p}$ were series expanded, it could be seen that (5) is an infinite order differential equation with constant coefficients, a hystero-differential equation. The transformed Eq. (8) is one of a type known as an "exponential sum." Reference is made to Langer[5] for a bibliography.

III. STABILITY

One approach,[4,6] which proved unsuccessful, stems from the argument that since in the open form of (7), the coefficients of the higher powers of s are very small, stability may be determined from a finite number of terms. That this procedure gives no information on stability may be seen by noting that if in a typical term of the expansion, say $c_n s^n$; $c_n>0$, then certainly $c_{n+1}<0$ for n sufficiently large. The closed sum would then give "possibly stable" or "unstable" results, depending on the number of terms used.

For the particular case of

$$Z(s)=as^2+bs+\beta s\epsilon^{-s}+c \tag{10}$$

Sherman[7] has found an analytical stability criterion that is particularly simple and useful. This criterion with suitable modifications might be applied to some other types of systems, but that it could be generally applicable is doubtful.

It has been suggested[8] that the Nyquist plot be used to determine stability. Chu[9] has made an extension of the root locus method to this case. Of course, no graphical procedure can be used with the facility of a criterion similar to Sherman's.

IV. A NEW APPROACH

If only one retarded term exists in a particular system, then the characteristic equation has the form

$$A(s_k)+K\epsilon^{-\tau s_k}B(s_k)=0, \tag{11}$$

where $s_k=\sigma_k+i\omega_k$, k takes on all positive integer values, τ is a constant, and K is a constant greater than zero.

Therefore,

$$K\epsilon^{-\tau s_k}=\frac{-A(s_k)}{B(s_k)}, \tag{12}$$

$$|K\epsilon^{-\tau s_k}|=K\epsilon^{-\sigma_k\tau}=\frac{|A(s_k)|}{|B(s_k)|}; \tag{13}$$

then substituting in (11)

$$A(s_k)+\frac{|A(s_k)|}{|B(s_k)|}\epsilon^{-i\omega_k\tau}B(s_k)=0 \tag{14}$$

or

$$\frac{A(s_k)}{|A(s_k)|}+\epsilon^{-i\omega_k\tau}\frac{B(s_k)}{|B(s_k)|}=0 \tag{15}$$

which implies that

$$\arg A(s_k)-\arg B(s_k)+\omega_k\tau=(2m+1)\pi;\qquad m=0,\pm1,\pm2,\cdots. \tag{16}$$

But from (12), it is seen that if $\tau\sigma_k<0$,

$$|A(s_k)|>K|B(s_k)|;\quad (k=1,2,\cdots). \tag{17}$$

Therefore the truth of (17) for each s_k is a necessary and sufficient condition for stability,† but is of no direct utility, since prior knowledge of the roots of (11) would be required.

Conditions (16) and (17), however, do suggest a

[4] L. A. Pipes, J. Appl. Phys. **19**, 617–623 (1948).
[5] R. E. Langer, Bull. Am. Math. Soc. **37**, 213–39 (1931).
[6] N. Minorsky, J. Franklin Inst. **232**, 519–551 (1941).
[7] S. Sherman, Quart. Appl. Math. **5**, 92–97 (1947).
[8] N. Minorsky, J. Appl. Mech. **64**, A65–A71 (1942).
[9] Yaohan Chu, Trans. Am. Inst. Elec. Engrs., Part II, 291 (November 1952).

† If $\tau<0$, the inequality in (17) would be reversed for stability, while if $K<0$, its sign may be absorbed into $-[A(s)/B(s)]$.

procedure for the approximate determination of any number of the roots of (11).

The function $K\epsilon^{-\tau s}$ maps the right half of the s plane into the interior of a circle centered at the origin and of radius K. Therefore, any s_k with positive real part would also be mapped into the circle's interior. If we consider $K\epsilon^{-\tau s}$ as a mapping function, the grid formed of constant σ and constant ω lines (where $s=\sigma+i\omega$) in the s plane will be transformed into concentric circles and radial lines respectively. In a sense, rectangular coordinates have been transformed into polar coordinates. Similarly, using the mapping function $-[A(s)/B(s)]$, the grid of constant σ and ω lines in the s plane is transformed into some curvilinear coordinate system. If the two new coordinate systems thus found are superimposed, a certain enumerable set of points, $\langle s_k\rangle$, which are the roots of (11), will be found that map into the same points by the use of either coordinate system. The set $\langle s_k\rangle$ may be easily found by constructing admissible σ and ω curves. The admissible ω curve is the locus of all points with the same value of ω as determined by either coordinate system. Similarly an admissible σ curve may be found. Every intersection of an admissible σ curve with an admissible ω curve must have the polar coordinates $(K\epsilon^{-\tau\sigma_k}, -\tau\omega_k)$ by the methods of construction used. Therefore, any number of zeros s_k of (11) may be found by this process. In addition, an unstable zero ($\sigma_k>0$) will show itself as an intersection inside the unit circle.

As an example we will use a characteristic equation that has been elsewhere considered,[9,10] so that a comparison of the different methods of solution can be made. Let the characteristic equation be

$$s^2+\epsilon^{-s}(s+0.3)=0. \tag{18}$$

In this case,

$$K=\tau=1 \quad \text{and} \quad -\frac{A(s)}{B(s)}=\frac{-s^2}{s+0.3}.$$

Since concentric circles and radial lines are already present on polar coordinate paper, the σ and ω contours of ϵ^{-s} need not be plotted. In Fig. 1, the solid lines are the constant σ and constant ω contours of $-[A(s)/B(s)]$ found by evaluating $-s^2/(s+0.3)$ along constant σ and ω lines in the s plane. The admissible ω curve is found by selecting a particular value of ω, say $\omega=1$ rad; then the intersection of the $\omega=1$ contour of $-s^2/(s+0.3)$ with the radial line at -1 rad $=57.3°$ will locate one point of the admissible ω curve. If τ were two instead of one as in the present example, a point of the admissible ω curve would be located at the intersection of the $\omega=1$ contour of $-s^2/(s+0.3)$ and the radial line at -2 rad. To construct the admissible σ curve, a value of σ is taken, say $\sigma=0$. A point of the admissible σ curve is then located at the intersection of the $\sigma=0$ contour of $-s^2/(s+0.3)$ with the circle of radius $1=\epsilon^0$. If K were equal to two

[10] Callender, Hartree, and Porter, (I) Trans. Roy. Soc. (London) **A235**, 415–445 (1936). (II) Proc. Roy. Soc. (London) **A161**, 460–480 (1936).

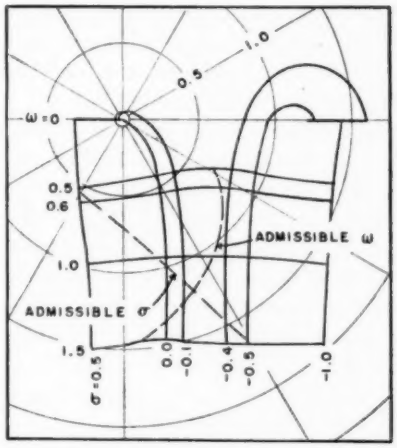


FIG. 1. Contours of σ and ω, admissible σ and admissible ω curves, for the characteristic equation $Z(s)=s+\epsilon^{-s}(s+0.3)$.

instead of one, the point would be at the intersection with the circle of radius $2=K\epsilon^0$. Evidently, the admissible σ and ω curves are very easily found and may be sketched in for several values of K and τ if desired. It is seen that the admissible σ and ω curves intersect at $s=-0.17+i1.2$ which compares favorably with the value $-0.1719+i1.1698$ previously obtained on a differential analyzer.[10] A real root‡ is also to be noted at $s=-0.4$ as compared with -0.4128 previously found.

The mapping is conformal almost everywhere; therefore the σ and ω contours are orthogonal. It might be noted that the $\omega=0$ contour is discontinuous, due to the pole of $-[A(s)/B(s)]$ at $s=-0.3$.

The σ and ω contours could be extended to obtain other roots, none of which could be unstable, since they must lie outside the unit circle. It is interesting to note that the other natural frequencies are very well approximated by adding multiples of 2π to the value $\omega=1.2$ found from Fig. 1. For example, $1.2+2\pi=7.48$

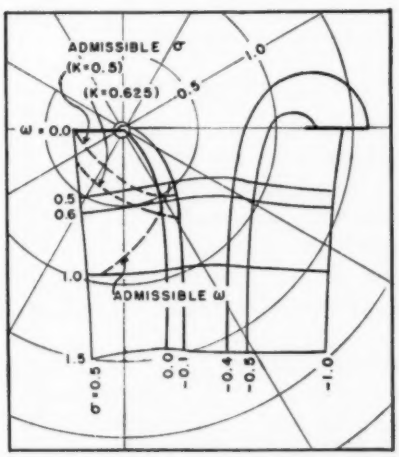


FIG. 2. Contours of σ and ω, admissible σ and admissible ω curves, for the characteristic equation $Z(s)=s^2+K\epsilon^{-1.75s}(s+0.3)$ where K takes on the values 0.5 and 0.625.

‡ If a more exact value of the real root is desired, it may be found by a reversion of (18), expressed in series form.

and $1.2+4\pi=13.76$ as compared with the values 7.5496 and 13.9280 obtained previously. It can be concluded that the roots of (18) are located in a fairly small sector of the fourth quadrant of Fig. 1.

A major advantage of this method is the ease with which K and τ may be varied to optimum values. A change in K or τ varies the position of the admissible σ or ω curve but leaves unchanged the contours of constant σ and ω. If the maximum value of K for stability is desired, it may be found by noting the point at which the admissible ω curve intersects the $\sigma=0.0$ contour of $-[A(s)/B(s)]$, which will be the limiting point for stability. In Fig. 1, this occurs at $K=1.33$.

Figure 2 has the same σ and ω contours as Fig. 1, but an admissible ω curve is shown for $\tau=1.75$, and two admissible σ curves, one for $K=0.5$ and the other for $K=0.625$, are shown. As seen from Fig. 2, the roots obtained are $-0.05+i0.6$ and $+0.05+i0.65$ which are evidently the dominant modes for the cases considered.

V. CONCLUSION

In many cases the segmented time function type of solution is too laborious to be useful.

The method given in the present paper and that of Chu[9] give simultaneously the stability and time solution of a given difference-differential equation, while the Nyquist plot yields only the stability. The methods of Chu and the present paper require about the same amount of labor initially; however, if a solution is desired for several values of K or τ, Chu's procedure requires the construction of a family of lines or the computation of a set of values as opposed to adding a single admissible σ or ω curve in the method of this paper. The addition of an admissible σ or ω curve requires practically no computation.

The author appreciates the advice of Professor V. C. Rideout of the University of Wisconsin and the financial support of The National Science Foundation.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 9 SEPTEMBER, 1954

Transient Response of a *p-n* Junction*

BENJAMIN LAX AND S. F. NEUSTADTER †
Lincoln Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received December 22, 1953)

The authors investigate the transient response of a *p-n* junction in a circuit where the applied voltage is suddenly switched from the forward to the reversed direction. The behavior is studied as a function of the circuit resistance. The solution of the appropriate diffusion equation is carried out for a time-dependent boundary condition, resulting in an integrodifferential equation. Analytical and numerical results for the current and the junction voltage as a function of time are then obtained by suitable approximations for the "recovery" and reverse periods. The "recovery" or reversal time t_f is the time required for the junction voltage to become negative and is given by

$$\operatorname{erf}\sqrt{\frac{t_f}{\tau}}=\frac{1}{1+(I_f/I_r)},$$

where τ is the lifetime of holes, I_f is the forward current, and I_r the initial reverse current, which is limited by the external resistance.

I. INTRODUCTION

A TYPICAL situation which is to be treated in this paper is shown in Fig. 1. A *p-n* junction is placed in a circuit, and a forward voltage V_f is applied until a steady-state current I_f is established. A switch S is suddenly thrown so that the forward voltage is removed and the junction is immediately placed in series with a resistance R and a voltage V in the reverse direction.

It is evident that the voltage V_J at the junction will not reverse immediately, because this would result in a very sharp carrier gradient at the junction and hence a large initial diffusion current in the reverse direction. The initial current in the reverse direction is limited to a finite value by the applied voltage V and the resistance R. Hence, the holes are removed from the n region at a finite rate and the junction potential must remain positive at the initial phase. Eventually, of course, the voltage across the junction will become negative. During the reverse phase, as the voltage becomes increasingly negative, a space-charge or transition layer is built up in which the free carriers are depleted.[1] This sets up a capacitance whose value is dependent upon the junction voltage V_J which appears across this transition layer. The current associated with the creation of the capacitance C, given by

$$I_c=C\frac{dV_J}{dt}+V_J\frac{dC}{dt},$$

* The research in this document was supported jointly by the Army, Navy, and Air Force under contract with the Massachusetts Institute of Technology.
† Staff Members, Lincoln Laboratory, Massachusetts Institute of Technology.

[1] W. Shockley, Bell System Tech. J. 28, 435 (1949).

is shown to be small compared to the diffusion current and can be usually neglected.

The analysis is made for a one-dimensional situation which assumes that the junction is formed between two semi-infinite slabs of p-type and n-type semiconductors. For simplicity, it is also assumed that the conductivity of the p-type material is much larger than that of the n type. It is then assumed that the entire current is carried by holes in the n region, which is long compared to a diffusion length L. Mathematically, the problem can be reduced essentially to an equivalent heat-flow problem in a semi-infinite slab whose surface is subjected to a time-varying temperature[2] (see Appendix I).

II. THEORY

A. Time-Dependent Solution

The hole density p in an n-type semiconductor for our one-dimensional case is governed by the diffusion equation[3]

$$\frac{\partial p}{\partial t}=D\frac{\partial^2 p}{\partial x^2}+\frac{p_n-p}{\tau}, \tag{1}$$

where p_n is the equilibrium hole concentration, τ the lifetime, and D the diffusion coefficient of the holes in the n region. The total current I of the diode due to diffusion can be specified as the hole current at the junction interface $x=0$, i.e.,

$$I=-qDA\frac{\partial p}{\partial x}\bigg|_{x=0}, \tag{2}$$

where q is the charge of the positive hole and A the area of the cross section of the diode. If the voltage across the junction interface is V_J, then the hole concentration p_0 at $x=0$ is expressed by the Boltzmann relation

$$p_0=p_n\exp\left[\frac{qV_J}{kT}\right],$$

where k is the Boltzmann constant and T is the absolute temperature.

In the steady-state forward position, $(\partial p/\partial t)=0$ and $V_J=V_0$, yielding the resulting expressions for the concentration and the current:

$$\begin{aligned} p&=p_n\left\{1+\left(\exp\left[\frac{qV_J}{kT}-1\right]\right)\exp\left[-\frac{x}{L}\right]\right\},\\ I_f&=-qDA\frac{\partial p}{\partial x}\bigg|_{x=0}=q\frac{DAp_n}{L}\left(\exp\left[\frac{qV_f}{kt}-1\right]\right), \end{aligned} \tag{3}$$

where $L=(D\tau)^{\frac{1}{2}}$.

[2] H. C. Carslaw and J. C. Jaeger, *Conduction of Heat in Solids* (Oxford University Press, London, 1947).

[3] W. Shockley, *Electrons and Holes in Semiconductors* (D. Van Nostrand Company, Inc., New York, 1951).

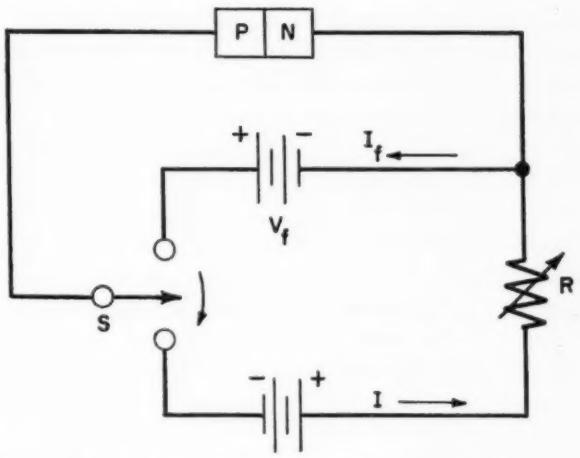


FIG. 1. Switching circuit for reversal of flow in a p-n junction.

When the voltage is reversed, the junction potential $V_J(t)$ becomes a function of time, varying from $V_J(0)=V_f$ to some limiting value $V_J(\infty)=V_l$, which is approached asymptotically after the reversal. The current through the junction is given by Ohm's law in terms of the reverse battery potential V, the series resistance R, and the junction potential, as follows:

$$V=V_J+IR. \tag{4}$$

The two unknowns are I and V_J. In order to solve for them, we need another relationship between them. This can be obtained from diffusion theory as follows: The concentration of holes that satisfies Eq. (1), that has the initial hole concentration of Eq. (3) at time $t=0$, and that is subject to the time-varying boundary condition of $p(0,t)=p_n\exp[qV_J(t)/kT]$, can be shown to be the following (see Appendix I):

$$\begin{aligned} p(x,t)-p_n=p_n&\frac{x\exp\left[-\frac{t}{\tau}\right]}{2(\pi D)^{\frac{1}{2}}}\int_0^t\left(\exp\left[\frac{qV_J(\lambda)}{kT}\right]-1\right)\\ &\times\exp\left[\frac{\lambda}{\tau}\right]\frac{\exp\left[-\frac{x^2}{4D(t-\lambda)}\right]}{(t-\lambda)^{\frac{1}{2}}}d\lambda\\ &+\frac{p_n\left(\exp\left[\frac{qV_J(0)}{kT}\right]-1\right)}{2(\pi Dt)^{\frac{1}{2}}}\\ &\times\exp\left[-\frac{t}{\tau}\right]\int_0^\infty\exp\left[-\frac{\mu}{L}\right]\\ &\times\left(\exp\left[-\frac{(x-\mu)^2}{4Dt}\right]-\exp\left[-\frac{(x+\mu)^2}{4Dt}\right]\right)d\mu, \end{aligned} \tag{5}$$

where the first term on the right-hand side of Eq. (5) gives, as a function of space and time, the concentration of holes injected into a semi-infinite slab by the voltage impressed at the junction interface. This represents a time-dependent source or sink of holes at the interface. The second term represents the decay of the initial steady-state distribution of holes given by Eq. (3), subject to the boundary condition that the concentration vanishes at $x=0$ for all time. Since we have a linear equation, by superposition the complete solution is the sum of the two terms. If Eq. (5) is differentiated and evaluated at $x=0$, from Eq. (2), the diode current is

$$I=qAD^{\frac{1}{2}}p_n\left\{\frac{\exp[-t/\tau]}{\sqrt{\pi}}\times\int_0^t\frac{G'(\lambda)d\lambda}{t-\lambda}+\frac{G(0)}{\tau}\left(1-\operatorname{erf}\sqrt{\frac{t}{\tau}}\right)\right\}, \quad (6)$$

where, for the sake of brevity, we have put

$$G(t)=\left(\exp\left[\frac{qV_J(t)}{kT}\right]-1\right)\exp[t/\tau], \quad (7)$$

and the prime denotes differentiation.

From Eq. (3), we can express the initial forward current I_f in terms of $G(0)$, and the amplitude of the saturation current I_s obtained for large reverse voltages by $I_f=I_sG(0)$, where $I_s=qA\sqrt{D}(p_n/\sqrt{\tau})$. Substituting into Eq. (6) and making use of some equalities derived in Appendix II we obtain

$$I=I_s\sqrt{\frac{\tau}{\pi}}\exp[-t/\tau]\times\int_0^t\left\{\frac{[G(\lambda)-G(0)\exp[\lambda/\tau][1-\operatorname{erf}(\lambda/\tau)^{\frac{1}{2}}]']}{(t-\lambda)^{\frac{1}{2}}}\right\}d\lambda, \quad (8)$$

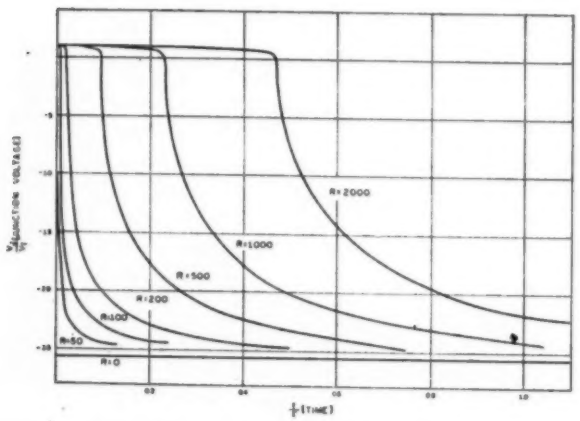


FIG. 2. Transient response of a junction diode potential for different values of series resistance.

which leads to

$$G(t)-G(0)\exp[t/\tau][1-\operatorname{erf}(t/\tau)^{\frac{1}{2}}]=\frac{1}{(\pi\tau)^{\frac{1}{2}}}\frac{1}{I_s}\int_0^t\frac{\exp[\lambda/\tau]I(\lambda)d\lambda}{(t-\lambda)^{\frac{1}{2}}}+K, \quad (9)$$

when we make use of the classical form of the solution to the Abel integral equation (see Appendix II). Since Eq. (9) also describes the current through the diode under steady-state forward conditions, i.e., $V_J=V_0$ and $G(t)=G(0)\exp[t/\tau]$, we conclude that $K=0$. It is desirable to normalize the time t by the characteristic lifetime τ of the material. Hence we shall put $r=t/\tau$, and $s=\lambda/\tau$. Also substituting for I from Eq. (4) into Eq. (9), the latter becomes

$$\left(\exp\left[\frac{qV_J(r)}{kT}\right]-1\right)I_s=I_f(1-\operatorname{erf}\sqrt{r})+I_R\operatorname{erf}\sqrt{r}-\frac{1}{R\sqrt{\pi}}\int_0^r\frac{\exp[-(r-s)]V_J(s)}{(r-s)^{\frac{1}{2}}}ds, \quad (10)$$

where $I_R=V/R$, which is a current in the reverse direction since V is negative.

B. Recovery Phase

During the recovery phase, $V_J(r)$ is positive, and, since $|V|\gg V_J$, the integral on the right side of Eq. (10) is small compared to the other terms, particularly to $I_R\operatorname{erf}\sqrt{r}$. There are then two possible choices for solving this equation, namely, we can ignore this term or else evaluate it by assuming that $V_J(s)$ has some average value $\bar{V}_J$. We shall choose to do the latter and shall demonstrate that the average value $\bar{V}_J$ is just a little less than the forward potential V_f, since the potential falls very slowly during the positive phase and drops abruptly at the end of this period. Hence no great error is incurred in assuming that $\bar{V}_J\approx V_f$ for the positive phase. With this assumption, Eq. (10) becomes

$$\left(\exp\left[\frac{qV_J}{kT}\right]-1\right)I_s=I_f+\left(I_R-I_f-\frac{V_f}{R}\right)\operatorname{erf}\sqrt{r}. \quad (11)$$

Since I_R is negative, we can substitute the positive quantity $I_r=V_f/R-I_R$. This represents the initial reverse current due to the reverse potential V and the forward potential V_f at the junction which appear across the series resistance R. Equation (11) can be solved for the junction potential V_J as a function of the normalized time r:

$$V_J=\frac{kT}{q}\log_e\left[1+\frac{I_f}{I_s}-\frac{I_f+I_r}{I_s}\operatorname{erf}\sqrt{r}\right]. \quad (12)$$

The first term in the parentheses can be neglected since $I_f\gg I_s$. Using this equation, we can also obtain the value of the period for the positive or forward phase by setting $V_J=0$. This period is designated as the "re-

covery" time r_f, and is the solution of the equation

$$\operatorname{erf}\sqrt{r_f}=\frac{I_f}{I_f+I_r}. \tag{13}$$

This equation is identical in form to the one obtained by Kingston[4] in his approximate solution of this problem, where he assumes that for the forward phase the diode current remains constant and is given by I_R. His solution does not take into account the contribution of the initial forward potential across the junction.

Using Eq. (12), V_J, normalized by V_f, was plotted as a function of r for $r\leqslant r_f$ and for different values of the series resistance R, as shown in Fig. 2. From Eq. (4), the diode current I, normalized by I_f, was also obtained for the forward period, and is shown in Figs. 3(a) and 3(b). It can be seen that the junction voltage during the recovery period, particularly for larger values of R, does not change appreciably from V_f except near $r=r_f$ where it drops very rapidly, thus justifying the approximation used to obtain Eq. (11).

C. Reverse Phase

The reverse phase is defined as the period for which $V_J(r)<0$. Since the time-dependent solution of the diffusion equation is given by Eqs. (8) and (10), these equations are the appropriate starting points for the solution during the reverse period. Since V_J is negative and $|V_J|\gg kT/q$ for $r>r_f$, except near $r=r_f$, and since $I_s\ll I_f$, the term on the left-hand side of Eq. (10) becomes negligibly small. Setting this term equal to zero, we obtain

$$I_f+(I_R-I_f)\operatorname{erf}\sqrt{r}=\frac{1}{R\sqrt{\pi}}\int_0^r\frac{\exp[-(r-s)]V_J(s)}{(r-s)^{\frac{1}{2}}}ds. \tag{14}$$

If we should assume that $V_J(s)\simeq V_f$ for $s\leqslant r_f$, then Eq. (14) would lead to

$$I_f+\frac{V_f}{R}\operatorname{erf}(r-r_f)^{\frac{1}{2}}-(I_r+I_f)\operatorname{erf}\sqrt{r}=\frac{1}{R\sqrt{\pi}}\int_{r_f}^r\frac{\exp[-(r-s)]V_J(s)}{(r-s)^{\frac{1}{2}}}ds, \tag{15}$$

which, after "formal" inversion, leads to

$$V_J(r)=\frac{\exp[-r]}{\pi}\frac{d}{dr}\int_{r_f}^r\frac{R\sqrt{\pi}\exp[\lambda][I_f+\frac{V_f}{R}\operatorname{erf}(\lambda-r_f)^{\frac{1}{2}}-(I_r+I_f)\operatorname{erf}\sqrt{\lambda}]}{(r-\lambda)^{\frac{1}{2}}}d\lambda, \tag{16}$$

for $r>r_f$.

This use of the inversion formula is not justified, nor does it lead to a correct result, since the term $V_f/R\times\operatorname{erf}(r-r_f)^{\frac{1}{2}}$ on the left-hand side of Eq. (15) does not have a bounded derivative at $r=r_f$. The correct value of $V_J(r)$ lies between the value given by Eq. (16), and that value obtained when the aforementioned term is neglected [which does not result in a large error since $|V_J(s)|$ is very small during the recovery phase compared to its value in the reverse phase]. As a matter of fact, when r is sufficiently larger than r_f, at least, say, $2r_f$, the error committed by taking $V_J(s)\simeq V_f$ for $s\leqslant r_f$ becomes negligible as one can readily see by examining the integrand in Eq. (14), for the contribution to the integral for s near r_f becomes negligible. With the preceding discussion in mind, we conclude from Eq. (16) that

$$V_J(r)\approx-\frac{2R}{\pi}(I_r+I_f)\arcsin\left(1-\frac{r_f}{r}\right)^{\frac{1}{2}}$$

for $r_f<r<2r_f<\frac{1}{5}$, and

$$V_J(r)=V_f-\frac{2R}{\pi}(I_r+I_f)\arcsin\left(1-\frac{r_f}{r}\right)^{\frac{1}{2}}, \tag{17}$$

for $2r_f<r<\frac{1}{5}$. {The upper bound on r is necessitated by the approximations $\operatorname{erf}\sqrt{r}\approx(2/\sqrt{\pi})\sqrt{r}$ and $\exp[\lambda]\approx 1$, $\lambda\leqslant r$, used in evaluating the integral in Eq. (16)}.

Another region where a useful approximation may be made is for $r\gg r_f$. Again we proceed from Eq. (16) which transforms into (see Appendix III):

$$V_J(r)=V_f+RI_f\left[\operatorname{erf}(r-r_f)^{\frac{1}{2}}+\frac{\exp[r_f-r]}{\sqrt{\pi}}\frac{1}{(r-r_f)^{\frac{1}{2}}}\right]-R(I_r+I_f)\frac{\exp[-r]}{\sqrt{\pi}}\frac{d}{dr}\times\left(\frac{4}{\sqrt{\pi}}\exp[r]\int_A\int\exp[-\rho^2]\rho d\rho d\theta\right), \tag{18}$$

where A is the shaded portion of the circle shown in Fig. 4. In many cases, r_f will be quite small; and, for r larger than r_f by a factor of 3 or more, the double integral may be taken over the entire quarter-circle. Equation (18) then becomes

$$V_J(r)=V-RI_f\left[1-\operatorname{erf}(r-r_f)^{\frac{1}{2}}-\frac{1}{\pi}\frac{1}{(r-r_f)^{\frac{1}{2}}}\exp(r_f-r)\right], \tag{19}$$

[4] R. H. Kingston, "Switching time in junction diodes and junction transistors," submitted to Proc. Inst. Radio Engrs. (August 5, 1953).

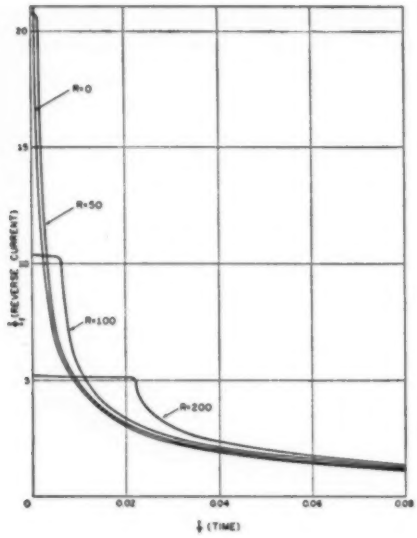


FIG. 3. The current response of a junction diode; (a) small values of series resistance; (b) higher values of a series resistance.

and hence the current is given by

$$\frac{I}{I_f}=1-\operatorname{erf}(\tau-\tau_f)^{\frac{1}{2}}-\frac{1}{(\pi)^{\frac{1}{2}}}\frac{1}{(\tau-\tau_f)^{\frac{1}{2}}}\exp[\tau_f-\tau]. \quad (20)$$

If we had not neglected the term I_s in the left-hand side of Eq. (10) to obtain Eq. (14), then Eq. (20) would read

$$I=I_f-(I_f+I_s)\left(\operatorname{erf}(\tau-\tau_f)^{\frac{1}{2}}+\frac{1}{\sqrt{\pi}}\frac{1}{(\tau-\tau_f)^{\frac{1}{2}}}\exp[\tau_f-\tau]\right), \quad (21)$$

which gives $\lim_{\tau\to\infty} I=-I_s$.

It is also of interest to note that Eq. (20) is the solution for the case $R=0$, except for the appearance of τ_f, which, of course, goes to zero. This zero-resistance solution is readily obtained from Eq. (6) by noting that $G(t)=-\exp[t/\tau]$, since the junction voltage becomes highly negative at the outset and it can be assumed that $V_J(\tau)=-\infty$. Then we find that:

$$I=I_f-(I_f+I_s)\left(\operatorname{erf}\sqrt{\tau}+\frac{\exp[-\tau]}{(\pi\tau)^{\frac{1}{2}}}\right). \quad (22)$$

This form of the limiting solution has been obtained by others[4,5] when I_s is neglected.

Without making the above approximations, we can obtain from Eq. (18) the more exact value of $V_J(\tau)$ for $\tau>\tau_f$.

$$V_J(\tau)=V-RI_f\left[1-\operatorname{erf}(\tau-\tau_f)^{\frac{1}{2}}-\frac{\exp[\tau_f-\tau]}{\sqrt{\pi}}\frac{1}{(\tau-\tau_f)^{\frac{1}{2}}}\right]$$
$$-R(I_r+I_f)\frac{2}{\pi}\exp[-\alpha]$$
$$\times\int_0^{\arctan(\tau_f/\alpha)^{\frac{1}{2}}}(\tan^2\theta)\exp[-\alpha\tan^2\theta]d\theta, \quad (23)$$

where $\alpha=\tau-\tau_f$. The last integral in Eq. (23) is then evaluated numerically by expanding it as a double series in τ_f and α.

D. Effect of Junction Capacitance

It was mentioned earlier that the current set up by the creation of a space-charge layer at the junction was small compared to the diffusion current when the voltage was reversed. We shall show this by considering the current due to the depletion of the free carriers at the junction surface. On the N side, this represents a flow of electrons in opposite direction to the flow of holes due to diffusion. Thus, the currents are additive and we can

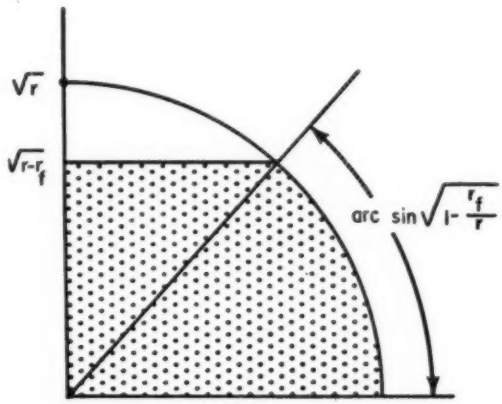


FIG. 4. Region of integration A.

[5] R. G. Shulman and M. E. McMahon, J. Appl. Phys. **24**, 1267 (1953).

consider that, in effect, a capacitance is placed in parallel with the diode, as shown in Fig. 5.

We then can write the following equations for the system:

$$V=(I_J+I_c)R+V_J,$$
$$I_c=C\frac{dV_J}{dt}+V_J\frac{dC}{dt}, \tag{24}$$

where V is the reverse potential applied to the diode; I_J represents the junction current due to diffusion as given in Eqs. (6) and (8); and I_c is the current due to the creation of the space-charge layer. The capacitance C set up by the latter can generally be specified in terms of V_J for a junction with a known impurity distribution.[1] Under these conditions, the two expressions in Eq. (24) can be reduced to a single integral differential equation which is more complex than the one considered previously, where the current due to the capacitance was neglected in order to solve the problem. Therefore, it remains only to show that this step is justified. This can be done by comparing the maximum value of I_c to the diffusion current at the same instant. This occurs when the slope of the junction potential V_J is at its peak, namely, when $\tau=\tau_f$ and $V_J=0$. At this time, $I_c=C\times(dV_J/dt)$. The value of this slope can be obtained from Eq. (12). Comparing this with the absolute value of I_J at $\tau=\tau_f$, we obtain

$$\frac{I_c}{I_J}=\frac{RC}{V\tau}\frac{kT}{q}\frac{I_f+I_r}{I_s}\frac{\exp[-\tau_f]}{(\pi\tau_f)^{\frac{1}{2}}}. \tag{25}$$

III. DISCUSSION OF RESULTS

The results of the theoretical development given in this paper are essentially summarized graphically in Figs. 2, 3, 6, and 7. The numerical work to obtain the graphs was made for a typical p-n junction with the following representative values of the essential parameters:

$D=44$ cm^2/sec, $A=0.0025$ cm^2,
$I_f=6\times10^{-3}$ amp, $p_n=2.5\times10^{12}$ cm^{-3},
$\tau=100$ μsec, $V=-6$ volts.

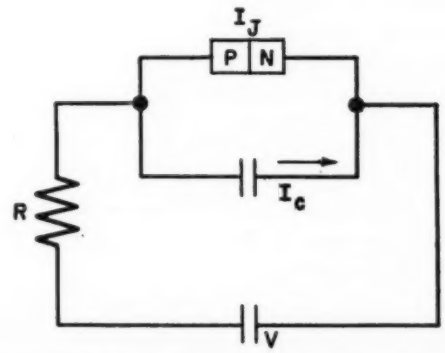


FIG. 5. Equivalent circuit of p-n junction and space-charge layer.

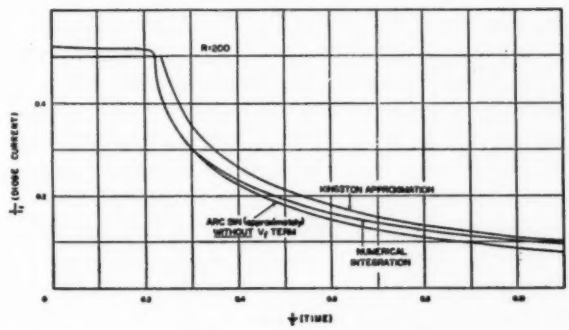


FIG. 6. Comparison of current response for a reversed junction diode as obtained by approximations to the more elaborate numerical method.

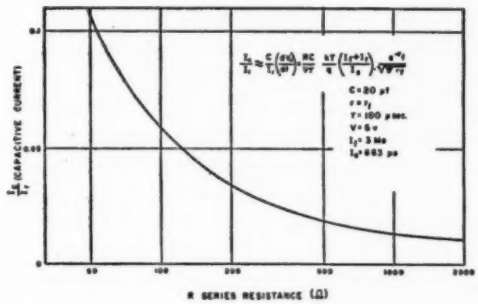


FIG. 7. Peak transient space-charge current in a reversed p-n junction as a function of series resistance.

Using the above values, the saturation current was calculated to be $I_s=6.63\times10^{-7}$ amp. For purposes of demonstrating the essential features of the theory, computations were carried out for values of the resistance R in series with the diode for 50, 100, 200, 500, 1000, and 2000 ohms.

The graph of Fig. 2 shows the variation of the junction voltage V_J normalized by the forward voltage V_f as a function of time normalized by the lifetime τ. The curves show that the recovery period increases with the resistance until it becomes comparable to the lifetime for values of the resistance of the order of 2000 ohms and greater. This represents, for our particular diode, a "hole storage" time of about 50 μsec. The curves also show a rapid decay of the voltage at the initial phase of the reverse period. For small resistances, the drop occurs in a brief instant and is almost equal to the total value of the applied voltage. For $R=0$, of course, the fall is instantaneous and reaches the limiting value V as represented by the horizontal line $V_J/V_f=-25.7$. For larger values of the resistance, the reverse phase is more gradual and approaches a limiting value V_s asymptotically.

Of equal interest are the curves of current I through the diode normalized by the forward current I_f. For small values of the resistance, the initial current through the diode is rather large and the decay is nearly that represented by the simple function of Eq. (21) for $R=0$. It is also interesting to note that, for practical

values of the series resistance, the current during the recovery phase diminishes slowly from its peak value. This justifies the assumption made by Kingston[4] that during this period the current remains constant and is essentially given by V/R. We then made a comparison of his result with that obtained for our more elaborate analysis. The comparison is shown in Fig. 6 for a representative situation. Kingston's approximation neglects the forward voltage V_f contributed by the stored holes at the junction for the recovery period. If this correction were made, the reverse phase would then be shifted even slightly to the left, giving a somewhat better agreement than that shown here. The arc sine approximation obtained from Eq. (17) is also compared with the solution obtained by the exact evaluation of the integral of Eq. (23). Needless to say, it may be concluded that Kingston's approximation is, for all practical purposes, quite satisfactory.

Finally, Fig. 7 shows the plot of the ratio of the capacitive current to the diffusion current as a function of the resistance for the particular diode under discussion. It can be seen that for practical values of the resistance this current is less than one-tenth of the peak reverse current I_r. Hence it can be neglected, particularly since the time constant associated with the creation of the space-charge layer is of the order of 10^{-9} sec. This represents a very small amount of charge as compared to that stored by the holes in the forward phase.

The authors are indebted to Mrs. Laura M. Roth for her contributions in carrying out the computations and for her helpful suggestions. We also wish to thank Dr. Kingston for many interesting and useful discussions of our work.

APPENDIX I. SOLUTION TO EQ. (1)

(A) If we let $z(x,t)=\exp[-t/\tau](p_n-p)$ for $t>0$ and $x>0$, Eq. (1) takes the form of the equation of heat flow

$$\frac{\partial z(x,t)}{\partial t}=\frac{D\partial^2 z(x,t)}{\partial x^2}. \tag{26}$$

Now, if $z(0,t)=0$ for $t>0$ and $z(x,0)=H(x)$ for $x>0$, then

$$z(x,t)=\frac{1}{2(\pi Dt)^{\frac{1}{2}}}\int_0^\infty H(\mu)\left(\exp\left[-\frac{(x-\mu)^2}{4Dt}\right]-\exp\left[-\frac{(x+\mu)^2}{4Dt}\right]\right)d\mu. \tag{27}$$

(B) If $z(0,t)=G(t)$ for $t>0$ and $z(x,0)=0$ for $x>0$, then

$$z(x,t)=\frac{x}{2(\pi D)^{\frac{1}{2}}}\int_0^t \frac{G(\lambda)\exp\left[-\frac{x^2}{4D(t-\lambda)}\right]}{(t-\lambda)^{\frac{3}{2}}}d\lambda. \tag{28}$$

(This result is readily obtained by means of Duhamel's theorem.[2])

APPENDIX II. TRANSFORMATION OF EQ. (6)

We note that the classical Abel integral equation

$$\int_0^t \frac{F(\lambda)d\lambda}{(t-\lambda)^{\frac{1}{2}}}=G(t),$$

has the solution[6]

$$F(\lambda)=\frac{1}{\pi}\frac{d}{d\lambda}\int_0^\lambda \frac{G(t)}{(\lambda-t)^{\frac{1}{2}}}dt. \tag{29}$$

Hence,

$$\int_0^t \frac{F(\lambda)d\lambda}{(t-\lambda)^{\frac{1}{2}}}=\exp[t/\tau]\left(1-\operatorname{erf}\sqrt{\frac{t}{\tau}}\right) \tag{30}$$

leads to

$$F(\lambda)=\frac{1}{\pi}\frac{d}{d\lambda}\left\{(\pi\tau)^{\frac{1}{2}}\cdot\exp(\lambda/\tau)\times\left[\operatorname{erf}\left(\frac{\lambda}{\tau}\right)^{\frac{1}{2}}-1+\exp(-\lambda/\tau)\right]\right\},$$

where use was made of the integral in Appendix III.

APPENDIX III. EVALUATION OF INTEGRALS

$$\int_{r_f}^r \frac{\exp\lambda}{(r-\lambda)^{\frac{1}{2}}}d\lambda=(\pi)^{\frac{1}{2}}\exp[r]\cdot\operatorname{erf}(r-r_f)^{\frac{1}{2}}, \tag{31}$$

and hence

$$\frac{d}{dr}\int_{r_f}^r \frac{\exp[\lambda]}{(r-\lambda)^{\frac{1}{2}}}d\lambda=\sqrt{\pi}\exp[r]\cdot\operatorname{erf}(r-r_f)^{\frac{1}{2}}+\exp[r_f]\frac{1}{(r-r_f)^{\frac{1}{2}}}.$$

$$\int_{r_f}^r \frac{\exp[\lambda]\operatorname{erf}\sqrt{\lambda}}{(r-\lambda)^{\frac{1}{2}}}d\lambda=\frac{4}{\sqrt{\pi}}\exp[r]\int_0^{(r-r_f)^{\frac{1}{2}}}\int_0^{(r-s^2)^{\frac{1}{2}}}\times\exp[-(s^2+z^2)]dzds, \tag{32}$$

$$=\frac{4}{\sqrt{\pi}}\exp[r]\int_A\int\exp[-\rho^2]\rho d\rho d\theta,$$

where $s^2=r-\lambda$, ρ, and θ, are the appropriate cylindrical coordinates for s and z, and A is the shaded portion of the circle shown in Fig. 4.

[6] E. J. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Macmillan Company, New York, 1945), p. 229.

Lower Modes of a Concentric Line Having a Helical Inner Conductor*

LOUIS STARK†
Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received October 26, 1953)

The natural, undamped modes of propagation in a concentric line having a circular outer conductor and a tape-helix inner conductor are investigated. The investigation is primarily concerned with a study of the propagation characteristics of the modes and, by making suitable approximations, closed expressions for the propagation factors of a few of the lower modes are obtained. Numerical results are then presented for a system having typical geometrical parameters, and generalizations of a qualitative nature are indicated. In the course of the work, the natural-mode field expressions are derived.

A. INTRODUCTION

HELICAL-CONDUCTOR transmission systems probably find greatest practical application as delay lines. Over a wide frequency range, with the exception of very low frequencies, a current wave travels at substantially the light velocity along the helical conductor, giving a relative retardation in the axial direction that is equal to the ratio of the length of a helix turn to the helix pitch. The unshielded helix cannot, however, propagate in this manner over an indefinite frequency range for, at a frequency such that adjacent helix turns would differ in phase by 2π radians, the system becomes a broadside antenna array. Of course, the same situation holds when the frequency would be such that neighboring turns were phased by 4π, 6π, etc. For a range of frequencies either side of the broadside frequencies, the radiation field would remain finite (being steered from endfire in one direction to endfire in the opposing direction as frequency is steadily changed) so that there is actually a series of frequency bands in which the open helix cannot be a wave-guiding system for a field whose current wave runs along the conductor at the light velocity. As frequency is increased towards a given band, the absence of a radiation field requires that the phase shift per turn decrease—a situation that proves to be impossible and results in a stop frequency band for the wave. Recent investigations of wave transmission along open helices[1,2] have shown quantitatively how the above phenomena occur. These investigations have shown another peculiarity of open helical systems that is important in delay-line applications, i.e., that the delay-line mode of propagation, which has been discussed qualitatively above and is presumably of greatest practical importance, is not a dominant mode of the system. Throughout the frequency range of its first pass band, two additional modes can also propagate.

It is probably true that shielded helical systems are used at least as often in delay-line applications as their unshielded counterparts. Though the helical element is common to both systems, they are, of course, fundamentally different in that the shielded system cannot lose energy by radiation. It therefore seems appropriate to investigate the propagation characteristics of the shielded system, and this is the primary aim of the work reported here. In the course of solving for the propagation factors of the natural modes, expressions for the fields in the line are derived.

The complexity of the system to be considered precludes the possibility of giving quantitative results for the general case. However, we will show such results for a particular set of geometrical parameters, and point out certain generalizations of a qualitative nature. In so doing, we hope to provide some results that will be useful in practical applications of the shielded helical system.

B. THE INDEPENDENT PARAMETERS FOR THE SYSTEM

We shall use as a model for the system a geometry of the type sketched in Fig. 1. As suggested by a previous analysis,[1] the helical conductor is taken to be an infinitesimally thin tape—an approximation that adapts the system to analysis in the circular-cylindrical coordinate system. The concentric line is characterized by the parameters: a=helix radius, p=helix pitch, δ=tape width, b=outer conductor radius, and the field frequency ω is specified by the number $k=\omega(\mu\epsilon)^{\frac{1}{2}}$ (μ and ϵ are assumed constant throughout the system). By the principle of similitude, one knows in advance that the solutions will depend only upon the linear dimensions

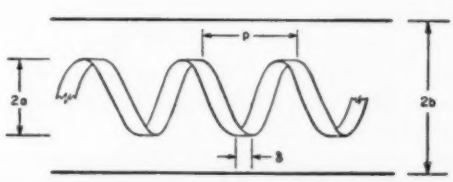


FIG. 1. The geometry of the helical line.

* The research in this document was supported jointly by the U. S. Army, Navy, and Air Force under contract with the Massachusetts Institute of Technology.
† Staff Member, Lincoln Laboratory, Massachusetts Institute of Technology.
[1] S. Sensiper, Electromagnetic Wave Propagation on Helical Conductors, Tech. Rept. No. 194, Research Laboratory of Electronics, Massachusetts Institute of Technology, May, 1951.
[2] V. J. Fowler, Helical Electrodes for Transmission Line Tubes, Electrical Engineering Research Laboratory Rept. No. 1-1, University of Illinois, Urbana, Illinois, March, 1953.

normalized with respect to wavelength, provided that system loss is neglected—as it will be. Thus, there are four independent parameters for the problem, and these are conveniently expressed by the set ka, b/a, $\psi=\cot^{-1}(2\pi a/p)$ and δ/p.

C. GENERAL APPROACH TO THE PROBLEM

As an approach to the problem, we shall use Floquet's theorem for periodic structures, which implies that we are considering an infinitely long structure that is source-free (otherwise the system is not truly periodic). This approach leads to the natural modes of propagation of the infinitely long system, and the modes so obtained are characteristic of a finite system if the modes are infinite in number and generate a complete set of field functions. No attempt will be made to prove that this requirement is fulfilled, but even with the lack of this proof one feels that the results should apply to the case of a finite structure if the terminals are far removed from the region of consideration.

D. GENERAL TECHNIQUE OF SOLUTION

When a natural mode propagates in the structure, a current will flow on the helical tape; this current will have a distribution that satisfies certain general requirements following from Floquet's theorem. Here we choose to begin by considering the conducting tape to be replaced by a fictitious tape of forcing current whose distribution satisfies these requirements but whose forcing propagation factor is arbitrary. The fields which arise from this tape of current within the outer pipe are then found, and finally, to obtain the natural-mode propagation factors, we indicate that the forcing current is in fact a natural-mode current by requiring tangential electric field to vanish at the tape surface.

E. THE NATURAL-MODE FIELDS

The distribution of the forcing current for the fields is not arbitrary but must be postulated as obeying the property that is known to be true of the natural-mode field quantities. This property is summarized in Floquet's theorem which, in this case, states that the linear current density $\bar{K}$ at the tape must obey the recurrence relation

$$\bar{K}(a,\theta,z+p)=\text{const}\times\bar{K}(a,\theta,z), \tag{1}$$

the constant being independent of the coordinates. The symmetry of the system suggests that along helical paths $[(2\pi/p)z-\theta]=$a constant; the current density should be a simple traveling wave and, assuming this to be the case, the vector components of $\bar{K}$ are of the form

$$K_\theta=f\left(\frac{2\pi}{p}z-\theta\right)\exp[-jh_0z], \tag{2a}$$

$$K_z=g\left(\frac{2\pi}{p}z-\theta\right)\exp[-jh_0z]. \tag{2b}$$

Upon noting that for single valuedness in θ the functions f and g must be periodic in their argument and of period 2π, one readily establishes that this assumed form is in accordance with Floquet's theorem. These expressions are not of the most general form which the theorem allows, for an additional periodic function of z, of period p, can be included as a factor. In view of the uniformity of the system along helical paths, this allowance for generality seems unwarranted. As a consequence, we will restrict the investigation to simple exponential variations along the tape. In the ensuing formal work, the propagation factor h_0 may be considered to be a complex quantity; however, since fields which are undamped in the z direction are of prominent interest, only real values are considered for subsequent numerical results.

The current density distribution across the tape determines the explicit forms of the functions f and g. Eventually, we will make some assumptions about this distribution, but for the time being we note only the periodic character of these functions. We now intend to expand K_θ and K_z in Fourier series and by this technique to represent the discontinuous driving current by a sum of continuous current sheets. The partial field (space harmonic field) arising from each of these sheets is then found, and the total field is obtained by summation.

Fourier series which represent K_θ and K_z in accordance with the requirements noted above are

$$K_\theta=\exp[-jh_0z]\sum_{m=-\infty}^{\infty}K_{\theta m}\exp\left[-jm\left(\frac{2\pi}{p}z-\theta\right)\right], \tag{3a}$$

$$K_z=\exp[-jh_0z]\sum_{m=-\infty}^{\infty}K_{zm}\exp\left[-jm\left(\frac{2\pi}{p}z-\theta\right)\right], \tag{3b}$$

where the Fourier coefficients $K_{\theta m}$ and K_{zm} are as yet general, but will follow from an assumed current distribution across the tape. Thus the mth component current sheet is

$$\bar{K}_m=(\bar{\imath}_zK_{zm}+\bar{\imath}_\theta K_{\theta m})\exp[jm\theta]\times\exp\left[-j\left(h_0+\frac{2\pi m}{p}\right)z\right], \tag{4}$$

and gives rise to partial electric and magnetic fields varying with θ and z as $\exp[jm\theta]\exp[-jh_mz]$, where $h_m=h_0+(2\pi m/p)$.

When the z dependence of the field is exponential, as it is here, the transverse field components follow from the longitudinal components E_{zm} and H_{zm} which are themselves readily obtained. The longitudinal components, being rectangular components, must satisfy the scalar wave equation (this follows from an elimination between the Maxwell equations), and from this it follows immediately that the r dependence of these

fields is a linear combination of Bessel functions of order m and argument

$$p_m r = (h_m^2 - k^2)^{\frac{1}{2}}$$

(for the modified Bessel functions). General solutions for the longitudinal components are thus clear. Constructing solutions for the two regions separated by the current sheet ($0 \lessgtr r < a$, $a < r \lessgtr b$) by adjusting the constants of combination in the general solutions so that

(1) E_{zm} and $\dfrac{\partial H_{zm}}{\partial r}$ vanish at $r=b$,‡

(2) E_{zm} and H_{zm} finite at $r=0$,

one obtains

$$E_{zm}{}^e = A_m\left[I_m(p_m r) - \frac{I_m(p_m b)}{K_m(p_m b)} K_m(p_m r)\right] \times \exp[jm\theta]\exp[-jh_m z], \quad (5a)$$

$$H_{zm}{}^e = B_m\left[I_m(p_m r) - \frac{I_m'(p_m b)}{K_m'(p_m b)} K_m(p_m r)\right] \times \exp[jm\theta]\exp[-jh_m z], \quad (5b)$$

$$E_{zm}{}^i = C_m I_m(p_m r)\exp[jm\theta]\exp[-jh_m z], \quad (5c)$$

$$H_{zm}{}^i = D_m I_m(p_m r)\exp[jm\theta]\exp[-jh_m z], \quad (5d)$$

for the partial longitudinal components internal and external to the surface $r=a$ to within the constants A_m, B_m, C_m and D_m that follow from the boundary conditions at the current sheet. The transverse components follow from the longitudinal components from a derived form of the Maxwell equations[3]

$$\bar{H}_{Tm} = j\frac{h_m}{p_m{}^2}\nabla_T H_{zm} + j\frac{k}{p_m{}^2}\left(\frac{\epsilon}{\mu}\right)^{\frac{1}{2}} \bar{i}_z \times \nabla_T E_{zm}, \quad (6a)$$

$$\bar{E}_{Tm} = j\frac{h_m}{p_m{}^2}\nabla_T E_{zm} - j\frac{k}{p_m{}^2}\left(\frac{\mu}{\epsilon}\right)^{\frac{1}{2}} \bar{i}_z \times \nabla_T H_{zm}, \quad (6b)$$

where, in the coordinate system used here,

$$\nabla_T = \bar{i}_r\frac{\partial}{\partial r} + \bar{i}_\theta\frac{1}{r}\frac{\partial}{\partial\theta}.$$

It is readily shown that the boundary condition $\bar{i}_r \times \bar{E}_{Tm} = 0$ at the outer pipe is automatically satisfied from the fact that E_{zm} and the normal derivative of H_{zm} have been required to vanish there.

‡ Although one begins by stating these homogeneous boundary conditions for the total field, the orthogonality in θ of the partial fields requires that the partial fields satisfy these requirements separately.

[3] R. B. Adler, Properties of Guided Waves on Inhomogeneous Cylindrical Structures, Tech. Rept. No. 102, Research Laboratory of Electronics, Massachusetts Institute of Technology, May, 1949.

One now has available all vector components of the partial field to within constant factors which are determined by the boundary conditions at the current sheet. To meet these conditions, it is sufficient to require that

$$\left.\begin{aligned} E_{zm}{}^i &= E_{zm}{}^e \\ E_{\theta m}{}^i &= E_{\theta m}{}^e \\ H_{zm}{}^i - H_{zm}{}^e &= K_{\theta m} \\ H_{\theta m}{}^e - H_{\theta m}{}^i &= K_{zm} \end{aligned}\right\} \text{ at } r=a.$$

One thus finds, after some algebra, that the mth partial field is given in terms of the corresponding current sheet excitation ($K_{\theta m}$ and K_{zm}) by the relations

$$E_{zm}{}^e = j\left(\frac{\mu}{\epsilon}\right)^{\frac{1}{2}}\frac{(p_m a)^2}{ka} I_m(p_m a) \times\left[\frac{K_m(p_m b)}{I_m(p_m b)} I_m(p_m r) - K_m(p_m r)\right] \times\left[\frac{mh_m a}{(p_m a)^2} K_{\theta m} - K_{zm}\right]\exp[jm\theta]\exp[-jh_m z], \quad (7a)$$

$$H_{zm}{}^e = (p_m a) I_m'(p_m a)\left[\frac{K_m'(p_m b)}{I_m'(p_m b)} I_m(p_m r) - K_m(p_m r)\right] \times K_{\theta m}\exp[jm\theta]\exp[-jh_m z], \quad (7b)$$

$$E_{zm}{}^i = j\left(\frac{\mu}{\epsilon}\right)^{\frac{1}{2}}\frac{(p_m a)^2}{ka} K_m(p_m a)\left[\frac{I_m(p_m a)}{I_m(p_m b)}\frac{K_m(p_m b)}{K_m(p_m a)} - 1\right] \times\left[\frac{mh_m a}{(p_m a)^2} K_{\theta m} - K_{zm}\right] I_m(p_m r) \times\exp[jm\theta]\exp[-jh_m z], \quad (7c)$$

$$H_{zm}{}^i = (p_m a) K_m'(p_m a)\left[\frac{I_m'(p_m a)}{I_m'(p_m b)}\frac{K_m'(p_m b)}{K_m'(p_m a)} - 1\right] \times K_{\theta m} I_m(p_m r)\exp[jm\theta]\exp[-jh_m z], \quad (7d)$$

and, by superposing all such partial fields, the representation for the total field is obtained. For brevity the expressions for the transverse fields have been omitted, but these are readily obtained from Eqs. (6a) and (6b).

F. FORMAL SOLUTION FOR THE NATURAL-MODE PROPAGATION FACTORS

Equations (7), together with Eqs. (6), give the partial fields which arise from a current distribution with an arbitrary propagation factor. At a given frequency, the propagation factor h_0 for a natural-mode current is, of course, not arbitrary, and it remains to specify that the tape current flows on a perfect conductor to obtain the natural-mode solutions; that is, one must require that the total tangential electric field at the tape, which will

be denoted by $\bar{E}_t$, vanish. Since the linear current density $\bar{K}$ vanishes everywhere off the tape, this boundary condition is conveniently stated by saying that the scalar product $\bar{E}_t \cdot \bar{K}^*$ is identically zero on the surface $r=a$. The reason for forming the product with $\bar{K}^*$, the complex conjugate of $\bar{K}$, rather than with $\bar{K}$ itself will be apparent later. Resolving the $\bar{E}_t$ and $\bar{K}^*$ vectors into components parallel ($\|$) and perpendicular ($\perp$) to the tape "winding" direction, this condition takes the form

$$E_{\|}K_{\|}^*+E_{\perp}K_{\perp}^*=0 \text{ at } r=a. \tag{8}$$

For an approximate solution, we will neglect the component of current density perpendicular to the tape direction, which is small for narrow tapes (tending to zero in the limit of vanishing tape width), so that the second term of Eq. (8) is ignored. Finally, we shall relax the requirement to the extent that only the integral of Eq. (8) over the surface $r=a$ must vanish; physically, this means that the natural-mode solutions are obtained by requiring that no net real or reactive power flows out of the tape surface. Computing $E_{\|}(a)=E_z(a)\sin\psi+E_\theta(a)\cos\psi$ from the field expressions, and noting that to within the approximation $K_\perp=0$,

$$\left.\begin{aligned} K_\theta &= K_{\|}\cos\psi \\ K_z &= K_{\|}\sin\psi \end{aligned}\right\} K_\perp=0, \tag{9}$$

the natural-mode specification takes the form

$$\int_0^{2\pi}\int_{-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\left\{K_{\|,m}F_m(p_ma,p_mb,h_ma,ka,\psi)\times\exp\left[-jm\left(\frac{2\pi}{p}z-\theta\right)\right]\right\}\sum_{n=-\infty}^{\infty}\left\{K_{\|,n}^*\times\exp\left[+jn\left(\frac{2\pi}{p}z-\theta\right)\right]\right\}dzd\theta=0, \tag{10a}$$

where the function F_m abbreviates a form which will presently be indicated explicitly. Assuming that the double sum can be integrated termwise, one sees from the orthogonality of the functions in θ and z that only the terms for which $m=n$ contribute, so that the result simplifies to

$$\sum_{m=-\infty}^{\infty}|K_{\|,m}|^2F_m=0, \tag{10b}$$

and now states that the complex powers carried by the separate partial fields into the surface $r=a$ must sum to zero. By substituting the explicit form of F_m in Eq. (10b), we obtain the final formal result

$$\sum_{m=-\infty}^{\infty}|K_{\|,m}|^2\left\{I_m(p_ma)K_m(p_ma)\left[1-\frac{K_m(p_mb)}{K_m(p_ma)}\frac{I_m(p_ma)}{I_m(p_mb)}\right]\times\left[\frac{p_ma}{ka\cot\psi}-\frac{mh_ma}{ka(p_ma)}\right]^2+I_m'(p_ma)K_m'(p_ma)\times\left[1-\frac{K_m'(p_mb)}{K_m'(p_ma)}\frac{I_m'(p_ma)}{I_m'(p_mb)}\right]\right\}=0. \tag{11}$$

To the extent that $K_\perp=0$, the characteristic Eq. (11) is an exact result which the natural-mode propagation factor h_0 (actually the product h_0a) must satisfy for given values of ka, $\cot\psi$, b/a, and δ/p. One sees that the result is, in a sense, incomplete, for there is no way of determining the weighting factors $|K_{\|,m}|^2$ that follow from the current distribution across the tape. However, one can assume a distribution which is at least qualitatively correct, and proceed on the assumption that the solutions are relatively insensitive to small departures from the correct distribution. Here we will assume from quasi-stationary considerations that the parallel component of current density goes infinite inversely as the square root of the distance from the tape edge, that it is symmetrical across the tape, and that the constant phase curves of the distribution are lines normal to the tape edge.§ For a current distribution of this type, $K_{\|,m}$, the Fourier coefficients of $K_{\|}$ in the series

$$K_{\|}=\exp[-jh_0z]\sum_{m=-\infty}^{\infty}K_{\|,m}\exp\left[-jm\left(\frac{2\pi}{p}z-\theta\right)\right]$$

are given by

$$K_{\|,m}=\frac{I}{p\cos\psi}J_0\left[\left(m+\frac{h_0a}{\cot\psi}\cos^2\psi\right)\frac{\pi\delta}{p}\right], \tag{12}$$

where I is the amplitude of the tape current, and J_0 is the Bessel function of zero order. When used in connection with the characteristic equation, the $K_{\|,m}$ will be written without the constant multiplier $I/p\cos\psi$.

G. APPROXIMATE FORMS OF THE CHARACTERISTIC EQUATION

Further approximation of the characteristic equation is necessary in order to obtain numerical results. Pertinent to the approximation procedure used here is a consideration of the axial phase velocities of the partial fields when the propagation characteristic falls in various regions of the ka, h_0a plane. The phase velocity

§ The current distribution is not, of course, uniquely specified by this description. The exact distribution we have in mind is considered in reference 1, where an evaluation of the Fourier coefficients may be found.

of the mth partial field v_{pm} is by definition

$$\frac{v_{pm}}{c}=\frac{k}{h_m}=\frac{ka}{h_0a+m\cot\psi}. \tag{13}$$

Following from Eq. (13), there is a simple graphical construction in the ka, h_0a plane that readily determines the phase velocity of any partial field for a given point on the propagation characteristic; this is stated simply as follows. The phase velocity of the mth partial field normalized with respect to the light velocity is given by the slope of the line drawn from the point $-m\cot\psi$ on the h_0a axis to the point on the characteristic. Figure 2 shows regions in the plane in which the various partial fields have phase velocity greater than c; in the quadrant subtended by the notation $\leftarrow m=0\rightarrow$, the 0th partial field has phase velocity greater than c, and similarly for the remaining quadrants. When the general mth partial field has phase velocity greater than c, its radial dependence in the guide has the oscillatory character of the I_m and K_m functions of imaginary argument. Solutions in these regions, which are here called the "exceptional regions," are therefore distinctly different in character from solutions outside.

It is apparent that for solutions lying in the strip of the plane $(s-\frac{1}{2})\cot\psi<h_0a<(s+\frac{1}{2})\cot\psi$, where s is some arbitrary integer which may include zero, the partial field with number $m=-s$ has the greatest magnitude of phase velocity throughout. By limiting attention to solutions which occur at values of ka well below the value $\frac{1}{2}\cot\psi$, one sees that the phase velocities of the remaining partial fields must be small as compared with c; as a consequence, these fields will be localized around the helical tape. Provided that b/a is a practical ratio, these fields will not be directly affected by the outer pipe. On the other hand, the partial field with number $m=-s$ can be strong over much of the waveguide cross section and can store a considerable fraction of the total field energy. One thus expects the term of the characteristic equation that is due to this partial field to be dominant in influencing the solution. It therefore seems natural to construct solutions for each s strip of the plane by separating out the dominant term of the characteristic equation for this strip (the term for $m=-s$) and balancing this term against the sum of the remaining terms. This is the procedure that will be used here. Throughout the sth strip, we shall approximate $K_{11,m}$ by $J_0[(m+s)\pi\delta/p]$ [see Eq. (12)] to facilitate the calculation of the remainder sums. To a second order and higher, the form of the remainder sums depends upon the strip being considered and, for the strips $s=0$ and $s=1$, in which we shall be primarily interested, a rather involved series of operations yields

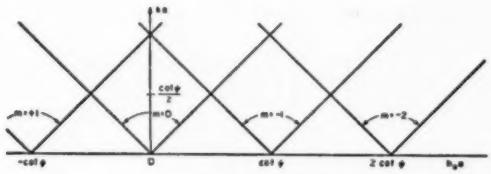


FIG. 2. Map of the "exceptional region" boundaries in the h_0a,ka plane.

for the sums

$$S_{R,0}=\frac{(h_0a)^2-(ka\csc\psi)^2}{(ka\cot\psi)^2}\sin\psi\left\{\log\frac{2}{\Delta}+\tfrac{1}{2}[(h_0a\sin\psi)^2(3\cos^2\psi-1)+(ka\sin\psi)^2]\times\left[1.202-\frac{\Delta^2}{2}\left(\frac{1}{2}+\log\frac{2}{\Delta}\right)\right]\right\}, \tag{14a}$$

$$S_{R,1}=\frac{(h_0a)^2-(ka\csc\psi)^2}{(ka\cot\psi)^2}\sin\psi\left\{\log\frac{2}{\Delta}+\tfrac{1}{2}[(h_0a\tan\psi-1)^2(3\cos^2\psi-1)\cos^2\psi+(k^2a^2-1)\sin^2\psi]\times\left[1.202-\frac{\Delta^2}{2}\left(\frac{1}{2}+\log\frac{2}{\Delta}\right)\right]\right\}, \tag{14b}$$

where $\Delta=\pi\delta/p$. These expressions comprise the leading terms of power series expansions and are correct to the second order in the terms inside the braces { }.

The form of the dominant term of Eq. (11) in a given s strip depends upon whether the field from which it arises travels with axial phase velocity greater or less than c and, as appearing in Eq. (11), the dominant term (along with every other term) is of the form which is appropriate when the corresponding field is a "slow wave." In the other case, where $h_m^2<k^2$ so that p_m is imaginary, one uses the relations for Bessel functions of imaginary argument and, letting $p_m=jq_m$, one obtains for the dominant term in this case

$$Q_m=\frac{\pi}{2}\left\{\frac{J_m(q_ma)}{J_m(q_mb)}[N_m(q_ma)J_m(q_mb)-N_m(q_mb)J_m(q_ma)]\times\left[\frac{q_ma}{ka\cot\psi}+\frac{mh_ma}{ka(q_ma)}\right]^2+\frac{J_m'(q_ma)}{J_m'(q_mb)}\times[N_m'(q_ma)J_m'(q_mb)-(N_m'(q_mb)J_m'(q_ma)]\right\}, \tag{15}$$

where $q_m=(k^2-h_m^2)^{\frac{1}{2}}$.

When the dominant field travels with phase velocity exactly equal to c, the two forms coalesce. We shall be particularly interested when this is the case for

the -1 partial field; when this occurs one has

$$Q_{-1}=-\frac{1}{2}\log\frac{b}{a}+\frac{1}{2}\left[1-\left(\frac{a}{b}\right)^2\right]\left[\frac{(h_0a)^2-(ka)^2}{(ka\cot\psi)^2}-\frac{1}{2}\right], \quad (16)$$

(when $ha=\cot\psi\pm ka$).

One can now write approximate forms of the solutions for the strips $s=0$ and $s=1$ which correspond to the ranges $-\frac{1}{2}\cot\psi<h_0a<\frac{1}{2}\cot\psi$ and $\frac{1}{2}\cot\psi<h_0a<\frac{3}{2}\cot\psi$, respectively.‖ Though we have given the remainder sums for each strip correct to the second order, one finds that it makes little difference in the final solutions if the second-order terms are neglected. To this approximation the remainder sums are the same for each strip and are given by

$$S_R'=\frac{(h_0a)^2-(ka\csc\psi)^2}{(ka\cot\psi)^2}\sin\psi\log\frac{2}{\Delta}$$

(correct to first order). (17)

Using this simplified form, the solutions indicated below are obtained.

I. $0<h_0a<\frac{1}{2}\cot\psi$

i. $ka<h_0a$

$$\left(\frac{p_0a}{ka}\right)^2=\cot^2\psi\frac{I_1(p_0a)K_1(p_0a)\left[1-\frac{K_1(p_0b)}{K_1(p_0a)}\frac{I_1(p_0a)}{I_1(p_0b)}\right]+(\sin\psi)\left(\log\frac{2}{\Delta}\right)}{I_0(p_0a)K_0(p_0a)\left[1-\frac{K_0(p_0b)}{K_0(p_0a)}\frac{I_0(p_0a)}{I_0(p_0b)}\right]+(\sin\psi)\left(\log\frac{2}{\Delta}\right)}, \quad (18a)$$

where $p_0=[h_0^2-k^2]^{\frac{1}{2}}$ and $\Delta=\pi\delta/p$. This is the solution for the familiar "delay-line mode" and, but for the term $(\sin\psi)(\log 2/\Delta)$ which brings in the effect of the tape width, it is identical to the solution that would be obtained if the helical sheath model were used for the tape helix. The sheath-model solution would replace this term by zero. It is clear that, in the case of small pitch angle helices, the two solutions should lead to nearly the same result. In the limit $p_0a\to 0$, which occurs as $ka\to 0$, one obtains from the limiting forms of the Bessel functions

$$\left(\frac{p_0a}{ka}\right)^2=\left(\frac{c}{v_0}\right)^2-1$$

$$=\cot^2\psi\frac{\frac{1}{2}\left[1-\left(\frac{a}{b}\right)^2\right]+(\sin\psi)\left(\log\frac{2}{\Delta}\right)}{\log\frac{b}{a}+(\sin\psi)\left(\log\frac{2}{\Delta}\right)} \text{ (as } p_0a\to 0), \quad (18b)$$

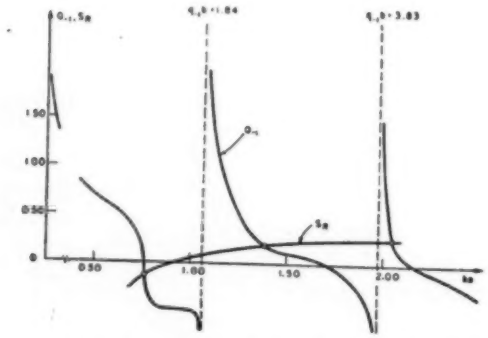


FIG. 3. A typical graphical solution for ka in the $s=1$ strip.

‖ It is indicated later that solutions for the right-half plane, $h_0a>0$, transform simply by a reflection about the ka axis to give all solutions in the left-half plane. It is thus unnecessary to formulate solutions for both halves of the $s=0$ strip.

which shows that for increased retardation at low frequencies the ratio b/a should be relatively small. As will be shown later in a typical numerical solution, the dispersion at low frequencies can be reduced considerably by such a choice.

ii. $ka>h_0a$

$$\frac{\pi}{2}\left\{\frac{J_0(q_0a)}{J_0(q_0b)}[N_0(q_0a)J_0(q_0b)-N_0(q_0b)J_0(q_0a)]\times\left[\frac{q_0a}{ka\cot\psi}\right]^2+\frac{J_1(q_0a)}{J_1(q_0b)}\times[N_1(q_0a)J_1(q_0b)-N_1(q_0b)J_1(q_0a)]\right\}$$

$$=\left[\left(\frac{q_0a}{ka\cot\psi}\right)^2+1\right](\sin\psi)\left(\log\frac{2}{\Delta}\right), \quad (19)$$

where $q_0=(k^2-h_0^2)^{\frac{1}{2}}$.

II. $\frac{1}{2}\cot\psi<h_0a<\frac{3}{2}\cot$

i. $|h_{-1}a|=|h_0a-\cot\psi|>ka$

$$I_1(p_{-1}a)K_1(p_{-1}a)\left[1-\frac{K_1(p_{-1}b)}{K_1(p_{-1}a)}\frac{I_1(p_{-1}a)}{I_1(p_{-1}b)}\right]\times\left[\frac{p_{-1}a}{ka\cot\psi}+\frac{h_{-1}a}{ka(p_{-1}a)}\right]^2+I_1'(p_{-1}a)K_1'(p_{-1}a)\times\left[1-\frac{K_1'(p_{-1}b)}{K_1'(p_{-1}a)}\frac{I_1'(p_{-1}a)}{I_1'(p_{-1}b)}\right]$$

$$=\frac{(ka\csc\psi)^2-(h_0a)^2}{(ka\cot\psi)^2}(\sin\psi)\left(\log\frac{2}{\Delta}\right), \quad (20)$$

where $p_{-1}=(h_{-1}^2-k^2)^{\frac{1}{2}}$.

ii. $|h_{-1}a|<ka$

$$\frac{\pi}{2}\left\{\frac{J_1(q_{-1}a)}{J_1(q_{-1}b)}[N_1(q_{-1}a)J_1(q_{-1}b)-N_1(q_{-1}b)J_1(q_{-1}a)]\right.$$
$$\times\left[\frac{q_{-1}a}{ka\cot\psi}-\frac{h_{-1}a}{ka(q_{-1}a)}\right]^2+\frac{J_1'(q_{-1}a)}{J_1'(q_{-1}b)}$$
$$\left.\times[N_1'(q_{-1}a)J_1'(q_{-1}b)-N_1'(q_{-1}b)J_1'(q_{-1}a)]\right\}$$
$$=\frac{(ka\csc\psi)^2-(h_0a)^2}{(ka\cot\psi)^2}(\sin\psi)\left(\log\frac{2}{\Delta}\right), \quad (21)$$

where $q_{-1}=(k^2-h_{-1}^2)^{\frac{1}{2}}$.

Numerical solutions in the strip $0<h_0a<\frac{1}{2}\cot\psi$ are obtained in a straightforward manner by assuming values of p_0a (or q_0a) and determining the value of ka which satisfies the equation. Low-frequency cutoffs occur at $h_0a=0$ for the modes which are solutions of Eq. (19); thus, to obtain the cut-off frequencies for

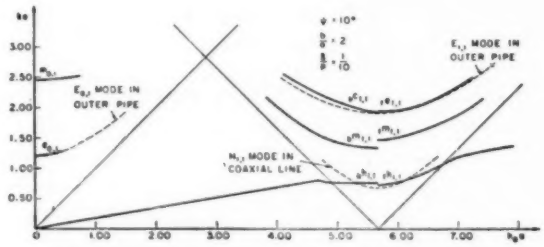


FIG. 4. Plot of the propagation characteristics.

these modes, one replaces q_0a by ka_{c0} throughout and solves for ka_{c0} (by a graphical procedure).

The solutions in the strip $\frac{1}{2}\cot\psi<h_0a<\frac{3}{2}\cot\psi$ can be obtained by a graphical procedure, plotting both sides of the pertinent equation as a function of ka for a fixed value of h_0a in this strip. A typical graphical solution is of the type sketched in Fig. 3. The poles of Q_{-1} occur at the zeros of the functions $J_1(q_{-1}b)$ and $J_1'(q_{-1}b)$; the zero of S_R' is, of course, at $ka=ha\sin\psi$. Solutions occurring exactly at the poles of Q_{-1} would indicate that the dominant -1 partial field distribution is that of a one-angular-variation E or H mode of the outer pipe, while those occurring at values of ka such that

$$N_1(q_{-1}a)J_1(q_{-1}b)-N_1(q_{-1}b)J_1(q_{-1}a)=0$$

or

$$N_1'(q_{-1}a)J_1'(q_{-1}b)-N_1'(q_{-1}b)J_1'(q_{-1}a)=0$$

(which are not critical points of Q_{-1}) would indicate that the dominant field is a one-angular-variation E or H mode of the corresponding uniform coaxial transmission line. As one might suspect, some of the modes can be regarded roughly as perturbations of these types; it will be seen presently that this suspicion is verified

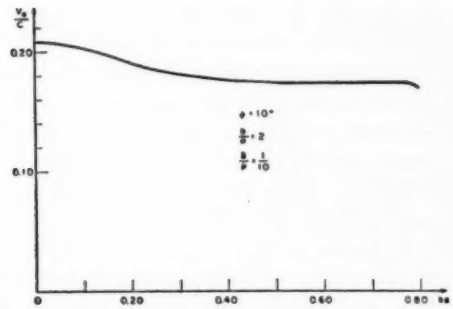


FIG. 5. Plot of the reciprocal retardation, or relative phase velocity, *vs* ka for the delay-line mode.

by numerical results. The extension of the above discussion to the consideration of solutions in the remaining strips follows in a rather obvious manner.

H. NUMERICAL RESULTS

The peculiarities of the solutions are best illustrated by numerical results for a typical geometry; in connection with these results, certain qualitative generalizations will be indicated. Solutions for the geometry having the parameters $\psi=10°$, $b/a=2$, $\delta/p=1/10$ are plotted in Fig. 4. In this plot, one recognizes the familiar low-frequency "delay-line mode" whose asymptotic retardation is given by $\csc\psi$, and notes that it undergoes a high-frequency cutoff at $ka\cong 0.80$, at which point the propagation characteristic is on the verge of entering the first "exceptional region." Essentially, this is also the cut-off condition for the mode when the system is unshielded. If, in general, the cutoff is determined by this condition, it is easy to show that

$$(ka)_{c0}=\frac{\cos\psi}{1+\sin\psi}\ \text{(for the delay-line mode)} \quad (22)$$

to within the approximation that the cutoff occurs exactly at the exceptional region boundary, and the relative retardation is equal to $\csc\psi$ at that point. The reciprocal retardation for this mode is indicated explicitly in the plot of Fig. 5, from which it is clear that the shielded helical transmission system is considerably less dispersive at low frequencies than its unshielded counterpart.

From purely physical considerations, one expects that a change of source location between the positions $z=\pm\infty$ results simply in a change of sign of the propagation factor h_0. This is implicit in the characteristic equation which determines only $|h_0|$; the sign to be affixed to the propagation factor is thus at one's disposal and can be chosen in accordance with the source position. If all modes are regarded as arising from a source at $z=-\infty$, then the slope of the propagation characteristic $d(ka)/d(h_0a)$, which is equal to the group velocity divided by c, must be positive for every mode. Thus, those characteristics having negative slope should actually be shown in a position which is reflected

about the ka axis. In the case of these modes, the current wave on the helix runs towards the source.

Of the solutions which lie in the plotted range of h_0a, the next higher mode is the $_fh_{1,1}$ mode, so called because this is a forward wave type and, in the vicinity of its low-frequency cutoff, it can be regarded as a perturbation of the $H_{1,1}$ mode in the corresponding uniform coaxial transmission line. By considering the solution for this mode in the general case, it can be shown that, so long as ψ is small, the near coincidence between the propagation characteristics at low frequencies for this mode and that of the $H_{1,1}$ mode is typical; one thus has a convenient way of locating approximately its lower cut-off frequency.¶ Although solutions in the remaining s strips have not been investigated quantitatively, it seems quite certain that solutions in these strips are generally higher modes than either the forward or backward wave branches of the $h_{1,1}$ mode.

Branches of the $m_{1,1}$ mode are not identifiable with one-angular-variation E or H modes of the outer pipe or the uniform coaxial transmission line, so that in this sense the $m_{1,1}$ mode is a mixed mode. The two lowest zero-angular-variation modes, excluding the delay-line mode, are the $e_{0,1}$ and $m_{0,1}$ modes; the dominant partial field of the latter is predominately neither E nor H in character. No notation is used to indicate whether a given mode is a perturbation of a uniform coaxial line mode or of a mode of the outer pipe, for to the extent that such modes can be regarded as perturbations it is true, in general, that h modes are perturbations of the uniform coaxial line modes, whereas e modes are perturbations of the outer pipe modes.

All modes having solutions at $h_0a=\cot\psi$, or equivalently, at $h_0p=2\pi$, are shown to have vanishing group velocity at that point, indicating that the total field is here a standing wave. This has been inferred from the numerical solution but not rigorously proven to be the case.

¶ Cut-off frequencies for the modes of coaxial transmission line are tabulated in reference 4.

[4] N. Marcuvitz, *Waveguide Handbook*, M.I.T. Radiation Laboratory Series, Vol. 10 (McGraw-Hill Book Company, Inc., New York, 1951).

I. INFLUENCE UPON MODE PATTERN OF CHANGES IN GEOMETRICAL PARAMETERS

Mode solutions for other values of the geometrical parameters have not been obtained numerically, but it is possible to state qualitatively how the mode pattern changes with these parameters. To this end, we shall consider separately the effects of changes in the ratio b/a and the pitch angle ψ; the effect of changing δ/p about practical values of the ratio is a small one and is ignored.

Changes in the ratio b/a affect the delay-line mode only at small values of ka where the effect is to reduce retardation with increasing b/a. The limiting retardation as $ka\rightarrow 0$ is readily calculated for any value of b/a from Eq. (18b). Modes whose fields are strong at the outer pipe—these are modes mapped in the "exceptional regions"—all occur at lower values of ka as the ratio b/a is increased; in regarding these modes roughly as perturbations of uniform wave-guide modes, they can be located approximately. In order that the "delay-line mode" propagate exclusively to some upper limit of ka, it is clear from the approximate lower cutoff condition of the $h_{1,1}$ mode what upper bound on the ratio b/a is allowable.

Changes in the helix pitch angle ψ, of course, influence the "delay-line mode" so that the asymptotic retardation is equal to $\csc\psi$; so long as the pitch angle remains small, it is true that to a first approximation the modes in the "exceptional regions" are not affected. The second-order effect of such changes is that decreasing pitch angle brings about a nearer coincidence between the lower cut-off frequencies of the helical system modes and the corresponding uniform wave-guide modes. This point is mentioned particularly with reference to the $h_{1,1}$ mode, whose lower cut-off frequency is of practical importance in designing for single-mode propagation.

ACKNOWLEDGMENTS

Sensiper's treatment of the unshielded helical coil has undoubtedly influenced the work reported here; the author is indebted to Dr. Sensiper for helpful discussions which acquainted the author with his work. Numerical calculations are due to Miss Evelyn Walker.

Minimum Noise Figure of Traveling-Wave Tubes with Uniform Helices

J. R. PIERCE AND W. E. DANIELSON
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received November 8, 1953)

By using a new theorem concerning noise in electron streams together with work by D. A. Watkins, an expression for the minimum noise figure of a traveling-wave tube with a uniform helix is deduced. For reasonable values of tube parameters this minimum lies around 6 db.

A THEOREM on the minimum noise in electron streams[1] makes it possible to compute minimum noise figures for various types of vacuum tubes.

Watkins[2] gives the noise figure of a traveling-wave tube with one noise source as

$$F=1+\tfrac{1}{2}(4-\pi)\frac{T_e}{T}\frac{r}{C}f(QC,d,\delta z), \tag{1}$$

where

$$f(QC,d,\delta z)=\left|(\delta_2+\delta_3)\cos\delta z-(\delta_2\delta_3-4QC)\frac{\sin\delta z}{(4QC)^{\frac{1}{2}}}\right|^2, \tag{2}$$

$$\delta=\frac{\omega_q}{u_0}=(4QC^3)^{\frac{1}{2}}\frac{\omega}{u_0} \tag{3}$$

and r is the ratio of the square of the velocity fluctuation at the maximum of the noise standing wave to that at the cathode

$$r=\frac{v_m^2}{v_a^2}. \tag{4}$$

It is proposed in reference 1 that there are two independent noise sources. It follows from the discussion there that if these are regarded as standing waves with maximum fluctuations in velocity v_{m1} and v_{m2}

$$\frac{v_{m1}^2v_{m2}^2}{v_a^4}=\frac{4\gamma^2(\omega_q/\omega)^2}{\sin^2\theta_d}=\frac{(4\gamma^2)(4QC^3)}{\sin^2\theta_d}. \tag{5}$$

Here γ is a factor which is about unity; it may be expressed in terms of a factor α used earlier[1] as

$$\gamma=\frac{\alpha}{4-\pi}=1.165\alpha. \tag{6}$$

θ_d measures the phase angle between the positions of the velocity maxima of the two noise standing waves; that is, if these lie Δz apart

$$\theta_d=\delta\Delta z. \tag{7}$$

Now, let the position of entrance of the electron stream into the helix be such that one standing wave has its velocity maximum $(2\theta+\theta_d)/2$ away from the entrance while the other standing wave has its velocity maximum $(2\theta-\theta_d)/2$ from the entrance. Watkins shows that $f(QC,d,\delta z)$ varies sinusoidally with δz, and we may write for the over-all noise figure, adding the uncorrelated noises from the two sources

$$F=1+\tfrac{1}{2}(4-\pi)\frac{T_c}{T}\frac{f_{\min}}{C(1-\beta)}\left(r(1-\beta\cos(2\theta+\theta_d))+\frac{1}{r}\frac{(4\gamma^2)(4QC^3)}{\sin^2\theta_d}(1-\beta\cos(2\theta-\theta_d))\right). \tag{8}$$

Here $f_{\min}$ is the minimum value, with respect to δz, of $f(QC,d,\delta z)$ and β is given by

$$\beta=\frac{f_{\max}-f_{\min}}{f_{\max}+f_{\min}}. \tag{9}$$

First we minimize F with respect to r and obtain

$$r=2\gamma(4QC^3)^{\frac{1}{2}}\left[\frac{(1-\beta\cos(2\theta-\theta_d))}{\sin^2\theta_d(1-\beta\cos(2\theta+\theta_d))}\right]^{\frac{1}{2}} \tag{10}$$

and for this value of r

$$F=1+4(4-\pi)\gamma\frac{T_c}{T}f_{\min}(QC)^{\frac{1}{2}}\times\left[\frac{(1-\beta\cos(2\theta-\theta_d))(1-\beta\cos(2\theta+\theta_d))}{(1-\beta)^2\sin^2\theta_d}\right]^{\frac{1}{2}}. \tag{11}$$

Equation (11) can be written

$$F=1+4(4-\pi)\gamma\frac{T_c}{T}f_{\min}(QC)^{\frac{1}{2}}\times\left[\frac{1-2\beta\cos2\theta\cos\theta_d-\beta^2+\beta^2\cos^22\theta+\beta^2\cos^2\theta_d}{(1-\beta)^2\sin^2\theta_d}\right]^{\frac{1}{2}}. \tag{11a}$$

By setting $\partial F/\partial\theta=0$ we obtain from (11-a)

$$\sin2\theta(\cos\theta_d-\beta\cos2\theta)=0. \tag{12}$$

If F is to have a minimum value, we must therefore have either

$$\cos2\theta=1 \tag{13}$$

or

$$\cos2\theta=\frac{\cos\theta_d}{\beta}. \tag{14}$$

[1] J. R. Pierce, J. Appl. Phys. **25**, 931 (1954).
[2] D. A. Watkins, Proc. Inst. Radio Engrs. **40**, 65–70 (1952).

Oddly enough, we will find that both (13) and (14) lead to the same final minimum. We should note, however, that β is less than unity and (14) cannot be satisfied for some values of θ_d. Let us proceed using (14), however. After substituting in (11), we obtain for the minimum noise figure

$$F_m=1+\gamma\frac{T_c}{T}4(4-\pi)(QC)^{\frac{1}{2}}f_{\min}\frac{(1-\beta^2)^{\frac{1}{2}}}{1-\beta}. \tag{15}$$

We see that this expression is independent of θ_d and must therefore hold for $\cos\theta_d=\beta$. As this is so, the minimum obtained when (13) is substituted into (11-a) and the resulting expression optimized with respect to θ_d must be that given by (15). Indeed, we find that this procedure leads to

$$\theta_d=\cos^{-1}\beta \tag{16}$$

as anticipated. It can be shown that when (13) holds, the maxima of the two uncorrelated standing-wave noise patterns are equal; that is

$$v_{m1}^2=v_{m2}^2. \tag{17}$$

When (13) does not hold and (14) applies, a range of values for the ratio v_{m1}^2/v_{m2}^2 is allowable by letting θ_d vary over the range specified by (14):

$$-\beta<\cos\theta_d<\beta. \tag{18}$$

We may note that, from (9),

$$\frac{f_{\max}}{f_{\min}}=\frac{1+\beta}{1-\beta}. \tag{19}$$

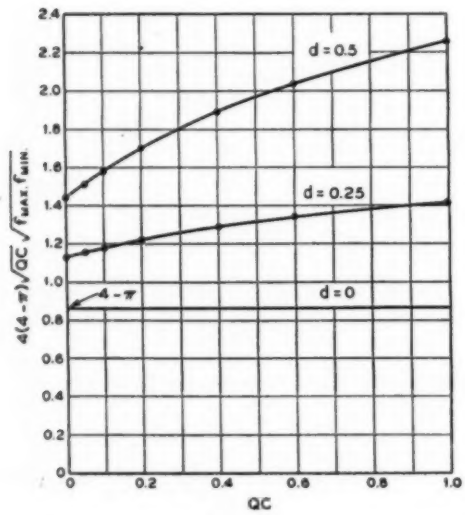


FIG. 1. The minimum noise figure is unity plus the product of γ, which is near to unity, times the ratio of cathode temperature to reference temperature, times the appropriate ordinate in the above graph. Curves are given for 3 values of the attenuation parameter d.

If we use (19) in connection with (15) we obtain

$$F_m=1+\gamma\frac{T_c}{T}4(4-\pi)(QC)^{\frac{1}{2}}(f_{\max}f_{\min})^{\frac{1}{2}}. \tag{20}$$

In Fig. 1, the quantity

$$4(4-\pi)(QC)^{\frac{1}{2}}(f_{\max}f_{\min})^{\frac{1}{2}}$$

has been plotted *vs* QC, under the condition that the tube is operated for maximum gain of the forward growing wave, for three values of the attenuation parameter d. The result for the case $d=0$ is particularly amenable to analysis, and the succeeding equations demonstrate that for this case

$$F_m=1+\gamma(4-\pi)\frac{T_c}{T}. \tag{21}$$

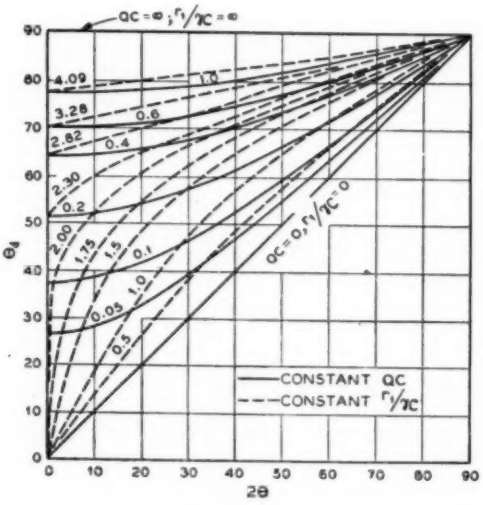


FIG. 2. Curves for the condition for minimum noise figure, showing the required θ_d, which measures the distance between the noise maxima from the two independent noise sources, plotted against 2θ, which measures the position of the input end of the helix with respect to the noise maxima.

To prove this we may express $(f_{\max}f_{\min})^{\frac{1}{2}}$ in terms of the real and imaginary parts of δ_2 and δ_3 as follows:

$$(4QC)^{\frac{1}{2}}(f_{\max}f_{\min})^{\frac{1}{2}}$$
$$=(4QC+y_2y_3)(y_2+y_3)+x_2^2y_3+x_3^2y_2. \tag{22}$$

Equation (22) is valid for $d\neq0$ as well as $d=0$. When $d=0$, the equations relating x_i and y_i are

$$x_i\{1+2y_i[(b+y_i)^2+x_i^2]\}=0, \tag{23}$$

$$x_i^2-3y_i^2-2by_i+4QC=0. \tag{24}$$

Here b is a velocity parameter introduced elsewhere.[3] If this parameter is chosen such that the increasing forward wave has maximum gain we have the additional

[3] J. R. Pierce, *Traveling-Wave Tubes* (D. Van Nostrand Company, Inc., New York, 1940), Chap. VII.

relation

$$y_1(3y_1^2+4y_1b+b^2-x_1^2)=0. \tag{25}$$

By choosing the appropriate roots of the three simultaneous Eqs. (23), (24), and (25), we find

$$b^3(1+4QCb)=(4QC)^3, \tag{a}$$

$$y_1=y_2=-\frac{4QC+b^2}{2b}, \tag{b}$$

$$x_1=-x_2=\left[\frac{1}{-2y_1}-(b+y_1)^2\right]^{\frac{1}{2}}, \tag{c}$$

(26)

$$y_3^2=\frac{1}{b+y_3}+4QC, \tag{d}$$

$$x_3=0. \tag{e}$$

If we now substitute $4QC/b$ for y_3 in Eq. (26-d) we can rearrange the terms to give (26-a). Hence

$$y_3=4QC/b \tag{27}$$

is the desired relation between y_3 and b. Using Eqs. (26) and (27) it is now easy to show that for $d=0$,

$$(y_2+y_3)(4QC+y_2y_3)+x_2^2y_3+x_3^2y_2\equiv\tfrac{1}{2} \tag{28}$$

which leads immediately to (21). Using Eqs. (26) and (27) it can also be verified that $(4QC)^{\frac{1}{2}}(f_{max}f_{min})^{\frac{1}{2}}$ has a minimum value with respect to changes in b for these same values of the roots. One may conclude, then, that when $d=0$, optimum gain and minimum noise figure are simultaneously obtainable. Numerical calculations show that this same conclusion is at least very nearly true for values of d in the range $0<d<1$.

The curves for $d=0.25$ and $d=0.50$ were obtained by use of Eq. (22); the requisite x_i's and y_i's (for b chosen to give maximum gain in the forward wave) were obtained by use of a generalization of (23) and (24) to include the cases $d\neq 0$.[4] It was found that points computed in this way were not in very good agreement with those obtained using Watkins' published data for f_{min} and f_{max}/f_{min}.[2] This discrepancy seems to result from inaccuracies in the published curves for f_{max}/f_{min}, particularly for small values of QC.

In the derivation of (20) certain relationships between θ_d, 2θ, and r have been specified as in Eqs. (10) and (14). Since β is a function of QC and d only, then when d is specified, a pair of values of θ_d and 2θ implies a definite value for QC. Also, by (5), (10), and (14) we obtain

$$\frac{r}{\gamma C}=\left(\frac{16QC}{1-\beta^2}\right)^{\frac{1}{2}}\left[2-2\left(\frac{\beta^2-\cos^2\theta_d}{\sin^2\theta_d}\right)^{\frac{1}{2}}-\frac{1-\beta^2}{\sin^2\theta_d}\right]^{\frac{1}{2}} \tag{29}$$

so that a definite value of $r/\gamma C$ is also implied. In Fig. 2 we are therefore able to plot curves of θ_d *vs* 2θ for constant values of QC (solid lines) and for constant values of $r/\gamma C$ (dashed lines) for a specific value of d (taken here as $d=0$). A plot of this type may be useful in deciding what choice of parameters θ_d, 2θ, QC, and $r/\gamma C$ (all of which give the same noise figure) are best in the sense that small changes in one of these parameters does not demand large changes in one or more of the others.

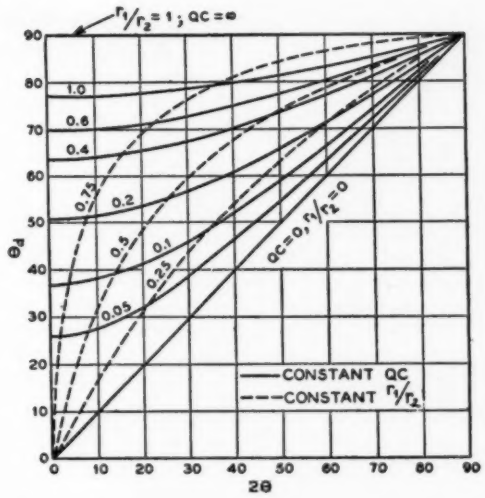


FIG. 3. Curves showing θ_d *vs* 2θ, holding constant a parameter different from that of Fig. 2.

Finally, it is of some interest to plot loci of v_{m1}^2/v_{m2}^2 = constant on a θ_d *vs* 2θ graph. Figure 3 is such a plot for the $d=0$ case, the dashed lines representing constant values of

$$\frac{v_{m1}^2}{v_{m2}^2}=\frac{r_1}{r_2}$$

and the solid lines representing constant values of QC.

What noise figure may we expect to attain in a traveling-wave tube on the basis of the foregoing analysis? Assume, for instance,

$$d=0,\quad \gamma\frac{T_e}{T}=3.5.$$

We obtain using Eq. (21)

$$F_m=4.0 \text{ or } 6.0 \text{ db}.$$

[4] See, for example, Eqs. (9) of reference 2.

Dielectric Constant and Loss Measurements on Barium Titanate Single Crystals While Traversing the Hysteresis Loop

M. E. Drougard, H. L. Funk, and D. R. Young
International Business Machines Corporation, Poughkeepsie, New York

(Received December 22, 1953)

A small amplitude high-frequency voltage has been superposed on a large amplitude low-frequency sweeping voltage to study the dielectric behavior of barium titanate single crystals while traversing the hysteresis loop. The output of a capacitance bridge (filtered to pass only the high-frequency component) is displayed on an oscilloscope as a function of either the applied sweeping voltage, the polarization, or the current to the sample. As a result, the bridge can be balanced for any portion of the sweeping voltage cycle desired. The dielectric constant and loss obtained in this manner have been studied as a function of measuring frequency, of sweeping frequency, of sweeping voltage and of rate of switching polarization (current). The measuring frequency and the switching current are shown to be the significant variables. The frequency dependence of the dielectric constant and loss can be described by a relaxation type of spectrum with a relaxation time of 5.5 microseconds.

I. INTRODUCTION

SINGLE crystals of barium titanate have been shown to exhibit relatively square hysteresis loops by von Hippel and co-workers at the Insulation Laboratory at M. I. T.,[1] and more recently by Merz at Bell Telephone Laboratories.[2] The sharp discontinuity in the hysteresis loop, for the best crystals, as the reversal of polarization begins suggests that an abrupt change in the characteristics of the dielectric must occur at this point. To learn more about the nature of this process, we were led to measure the dielectric constant and loss for small fields as the hysteresis loop is traversed. The behavior was found to be entirely different for the region of saturation and for the switching region, as was expected.

The frequency dependence of the initial dielectric constant for polycrystalline barium titanate has been shown to follow a relaxation type of spectrum[1] with a relaxation frequency of about 100 megacycles. Corresponding measurements are not available for single crystals, however, it is known that the relaxation frequency also lies above 10 mc, if measured in the "*c*" direction and when the crystal contains only a single domain. It has been known for some time that the times associated with the polarization switching phenomenon are much longer. Our measurements show that our crystals exhibit a relaxation spectrum while switching occurs, with a relaxation time of 5.5 microseconds corresponding to a relaxation frequency of 29 kc.

Using ceramic forms of barium titanate, Roberts[3] observed a large increase in the loss component measured during the steep portion of the loop. By superposing high-frequency parasitic loops on a low-frequency sweeping voltage, Jonker and Van Santen[4] observed a differential dielectric constant corresponding to the slope of the saturated portion of the loop (Fig. 1). As can be seen for polycrystalline barium titanate, the hysteresis loop is not sufficiently square to yield two well-defined regions as is the case for single crystals.

The crystals used in our measurements were grown in our laboratory and the *c* axis was made to be perpendicular to the face of the crystals by means of a high-field poling technique. We have obtained similar results from crystals grown by J. Remeika at the Bell Telephone Laboratories.

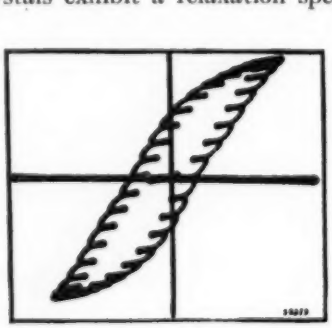


FIG. 1. *D vs E* for polycrystalline barium titanate with parasitic loops superposed on main hysteresis loop.

II. EXPERIMENTAL ARRANGEMENT

The circuit arrangement used for our experiments, is shown in Fig. 2. The crystal under study is cycled through its hysteresis loop by means of a low-frequency

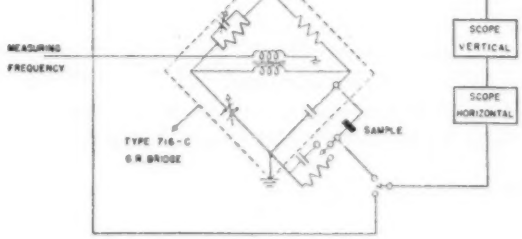


FIG. 2. Bridge, sweep, and display circuit.

[1] A. von Hippel, Revs. Modern Phys. **22**, 221 (1950).
[2] W. Merz, Phys. Rev. **91**, 513 (1953).
[3] S. Roberts, Massachusetts Institute of Technology (Doctoral thesis).
[4] Jonker and Van Santen, Philips Tech. Rev. **2**, 183 (1949).

switching voltage of 10 to 200 cycles per second supplied by a low output impedance (1000 ohms) amplifier.

A high-frequency sinusoidal voltage, called the measuring voltage, is superposed on the switching sine wave. In order not to interfere with the switching of the crystals, the high-frequency voltage is kept several hundred times less than the coercive voltage for the sample being studied. We have observed no effect on the hysteresis loop as a result of the high-frequency voltage.

The switching voltage is applied directly to the detector terminals of a GR type 716-C capacitance bridge and thus appears across the crystal, which is plugged in the "unknown" terminals. The measuring voltage is applied to the generator terminals and the mixing is accomplished in the bridge input transformer. The bridge unbalance appears across the detector terminals together with the switching voltage. The resulting voltage is fed into a high-pass filter so that its component corresponding to the measuring frequency only is available at the input of a wide-band, high gain amplifier, the output of which is fed to the vertical axis of an oscilloscope, through a sharply tuned filter.

A small resistor or a large capacitor can be placed between one electrode of the crystal under study and ground. The voltage appearing across either of them, which is proportional to the current flowing through the sample or the polarization of the crystal respectively, is fed to an amplifier and applied to the x axis of the oscilloscope.

The unbalance voltage of the bridge can thus be displayed on the oscilloscope against either the switching voltage E, the polarization P of the crystal, or the current I flowing through the crystal. The patterns thus displayed on the scope give no more than an indication of the variation of the impedance offered by the crystal to the measuring frequency. However, the bridge can be balanced and this complex impedance determined accurately for any desired value of either E, P, or I. A corrective term has to be applied to the values obtained in order to account for the impedance which we have placed in series with the crystal. For instance, if C' and G' are the values of capacitance and parallel conductance determined from the measurements and if R is the resistance (small) between the sample and ground, the corrected values are:

$$C=C'(1+2RG'), \quad G=G'(1+RG').$$

Under good operating conditions, the accuracy matches that of the capacitance bridge. The results obtained were found to be independent of the amplitude of the high-frequency voltage applied, if this was kept small with respect to the coercive voltage.

III. RESULTS

As shown in Fig. 3, the high-frequency impedance of the crystal changes abruptly, during the switching time,

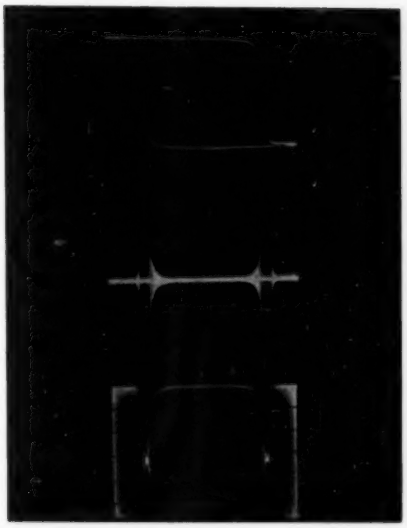


FIG. 3. Bridge output as a function of applied voltage.

from its value at saturation. It is interesting to note that the bridge balance is maintained uniformly over the entire region of saturation. These measurements indicated that the maximum unbalance was strongly dependent on both the amplitude and frequency of the switching sine wave, and was an increasing function of both.

A more detailed picture of the phenomena is obtained by using the polarization as the independent variable. As can be seen from Fig. 4, the maximum of bridge unbalance does not occur for zero polarization but is slightly offset.

It was found experimentally that the current was the significant variable. By means of the presentation shown in Fig. 5. the variation of the impedance with amplitude of the switching voltage was shown to be due

FIG. 4. Bridge output as a function of polarization.

Switching current pulse
Bridge balanced at saturation
Bridge balanced for maximum of switching current
Switching current

FIG. 5. Bridge output as a function of current.

to the fact that as this amplitude was increased, switching occurred at a steeper portion of the applied sine wave, which is similar to an increase in frequency at at constant applied voltage. The peak current increases if we increase the sweeping frequency, but can be brought back to the same value by decreasing the amplitude of sweeping voltage, and we find then that the corresponding impedance of the sample returns to its original value. The maximum variation of the impedance does not occur at zero polarization because the maximum of the switching current for this crystal occurs earlier in the cycle.

The crystals available were about 0.1 mm thick and the area of the electrodes about 0.5 mm.^{2} In order to keep both capacitance and dissipation factor of such crystals within the limits of the bridge, the switching current density had to be kept below 500 amp/m^2. In addition, in order to achieve a good discrimination between the high-frequency harmonics of the switching

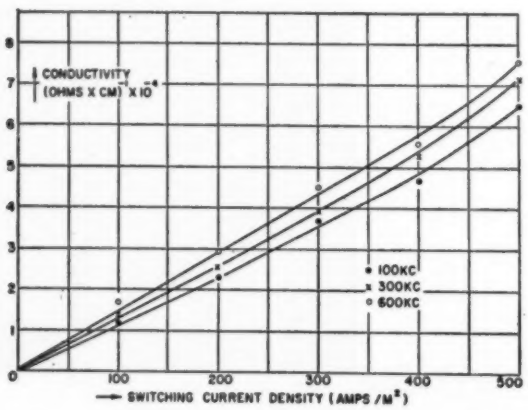


FIG. 6. Conductivity *vs* switching current density.

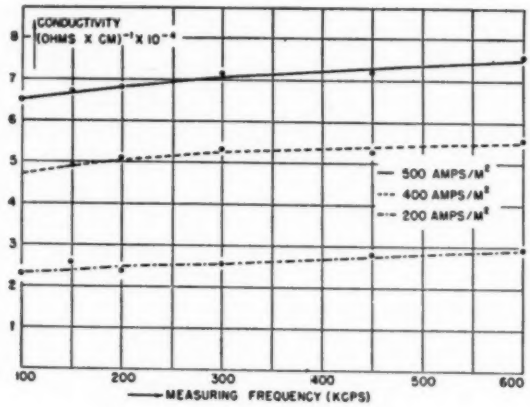


FIG. 7. Conductivity *vs* measuring frequency.

current and the measuring frequency, and due to the limitations of the bridge, this measuring frequency was limited to the 100 to 600 kc range.

The most striking result is a very large increase of the loss component of the high-frequency current flowing through the crystal while switching. This is similar to the earlier measurements of Roberts.[3] It has been plotted in terms of the equivalent conductivity σ (Fig. 6), related to k'' and tanδ by the formula

$$k^*=k'+jk''=k'(1+j\tan\delta),$$

$$\sigma=\omega k'\epsilon_0, \quad \tan\delta=\omega\epsilon', \quad \tan\delta=\omega k''\epsilon_0.$$

For the crystals under study, this equivalent conductivity for the peak switching current increases with an increase in frequency (Fig. 7).

The real part of the complex dielectric constant k' for a constant value of switching current shows a strong frequency dependence quite consistent for the crystals studied and decreases markedly as the measuring fre-

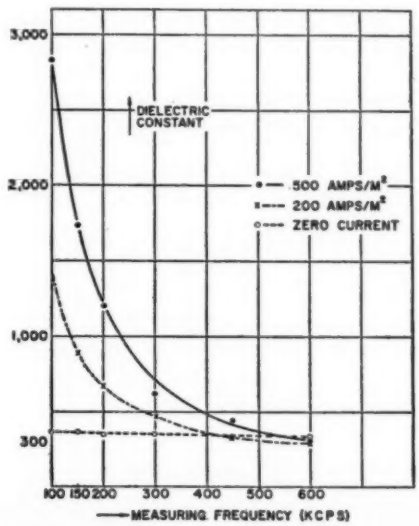


FIG. 8. Dielectric constant *vs* measuring frequency.

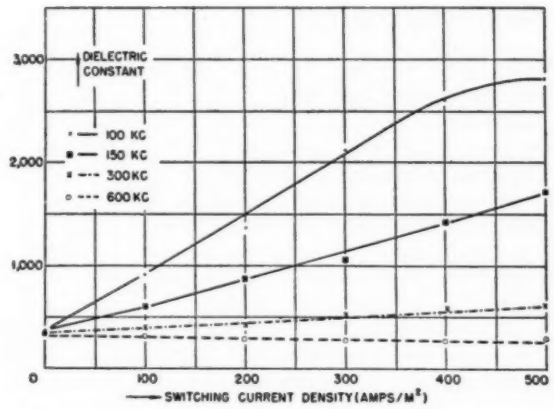


FIG. 9. Dielectric constant *vs* switching current density.

quency is increased (Fig. 8). The dielectric constant k' also increases with switching current (Fig. 9) for low measuring frequencies but is nearly constant for high measuring frequencies.

DISCUSSION OF RESULTS

This behavior suggested that our data might fit a Debye relaxation type of spectrum as follows:

$$k^*=\frac{\epsilon^*}{\epsilon_0}=\frac{a}{1+j\omega}+k_\infty',$$

where ϵ^* is the complex dielectric constant and τ is the relaxation time.

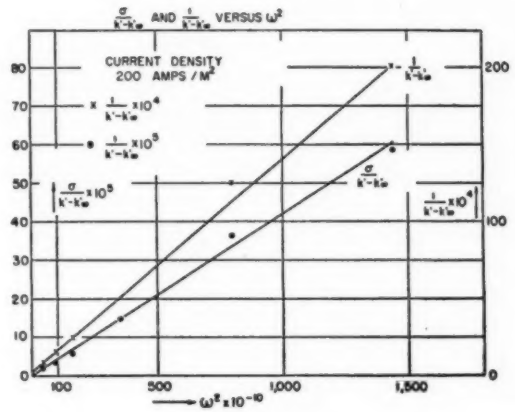


FIG. 10. $\sigma/k'-k_\infty'$ and $1/k'-k_\infty'$ *vs* ω^2.

From this equation we find that k' and σ depend on the frequency in the following manner:

$$\frac{1}{k'-k_\infty'}=\frac{1+\omega^2\tau^2}{a},$$

$$\frac{\sigma}{k'-k_\infty'}=\omega^2\tau\epsilon_0.$$

Plots of $1/(k'-k'_\infty)$ or of $\sigma/k'-k_\infty'$ *versus* ω^2 should reduce to straight lines, the slope being respectively τ^2/a and $\epsilon_0\tau$. This is verified in Fig. 10.

The values for τ and k_∞' that had to be used to make the results fit the above equations are given in Table I.

The quantity "a" is proportional to the switching current density J with the proportionality constant $b=98$ m²/amp. This is consistent with the linear dependence of k' and σ on the switching current density shown by Figs. 6 and 9.

TABLE I.

Amp/m²	100	200	300	400	500
τ(μsec)	5.2	5.4	5.5	5.5	5.5
k_∞'	290	250	210	215	230

Newton, Ahearn, and McKay[5] observed the time duration of current pulses in barium titanate crystals that they interpreted to be due to the reversal of individual domains. The times they observed were from 1–10 microseconds which is in agreement with our results.

It should be emphasized that the relaxation time we have measured is for an applied field equal to the coercive field. If higher fields are applied, switching can occur in a shorter time as recent results of Merz[6] indicate.

V. ACKNOWLEDGMENT

The crystals used were grown by C. Karan and B. Skinner. Most of the data were taken by E. Davin. We are indebted to Dr. S. O. Morgan of Bell Telephone Laboratories for the opportunity to confirm our results on their very nice crystals. Dr. G. A. Jonker kindly furnished us with the negative from which Fig. 1 was taken.

[5] Newton, Ahearn, and McKay, Phys. Rev. **75**, 103 (1949).
[6] W. Merz (private communication).

Solute Redistribution by Recrystallization

ROBERT G. POHL
Research Department, The Rauland Corporation, Chicago 41, Illinois
(Received January 4, 1954)

The time-dependent diffusion equation is solved using transformation calculus to obtain the solute concentration in a solvent after a segregating liquid-solid interface is passed through the material. The results are presented in graphical and tabular form for typical experimental conditions and compared with previously published findings.

INTRODUCTION

THE distribution of impurities in a melt of impure metal or alloy is altered by the solidification of a crystal from the melt. The ratio of the equilibrium concentration of impurities or solute in the solid to the concentration in the liquid *immediately adjacent* to the solid is generally considered to be a constant for any particular solute in any particular solvent[1,2] at some given temperature and is defined as the segregation constant k.

Pfann[3] has calculated the solute concentration in the solid assuming complete mixing in the melt for both normal freezing, wherein the whole boat is molten initially and slowly recrystallized from one end, and for zone melting, wherein several molten zones are passed through the material. The resulting equation is velocity independent, in contradiction with experiment, with the distance dependence of the solute distribution in the solidified crystal essentially determined by the length of the boat.

Tiller[4] *et al.* have recently considered the effects of diffusion in the liquid, assuming these to predominate over those of mixing and convection, obtaining the relation for normal freezing

$$C_s=C_0\left\{k+(1-k)\left[1-e^{-k\frac{R}{D}x}\right]\right\}, \tag{1}$$

where C_0 is the initial concentration in the melt and is assumed to be independent of position, D is the diffusion constant for the solute in liquid solvent at the melting point of the solvent, R is the velocity at which the liquid-solid interface advances and is assumed to be constant, x is the distance measured from the end of the boat initially recrystallized, and C_s is the solute concentration in the solid as a function of distance for an infinitely long boat of constant cross section.

Equation (1) was derived by Tiller *et al.*, assuming "that the rate of approach of C_s to C_0 with distance along the specimen is proportional to (C_0-C_s) at any distance x measured from the beginning of the specimen," yielding the exponential decay. This paper will derive an equation similar, but not equal, to (1) by considering the time dependent diffusion equation and the boundary conditions at the liquid-solid interface. This result can then be utilized in an iterative manner to obtain the solution for a second pass through an already segregated crystal. The method of solution can be applied to effects of changing the rate at which the interface advances, and to boundary layer effects in actual systems. The solution proposed by Burton, Prim, and Slichter[5] to the transport problem wherein the diffusion dominated region is limited by boundary layer flow will be considered further on in this paper.

DIFFERENTIAL EQUATION

The diffusion of solute in liquid solvent will be assumed to be governed by the time dependent diffusion equation

$$\frac{\partial c}{\partial t}=D\frac{\partial^2 c}{\partial x_b^2}, \tag{2}$$

where $c=c(x_b,t)$ is the solute concentration in the liquid at a time t at a point x_b; x_b being a coordinate measured with respect to the fixed boat. Further, we shall assume 1. Mass transport by recrystallization is negligible. 2. The segregation constant k is constant. 3. Turbulence is negligible. 4. Diffusion in the solid is negligible. 5. Cross-sectional area is constant. 6. Vaporization is negligible.

The solution of the problem is simplified by establishing a coordinate system fixed with respect to the moving interface. Let x=distance from interface of a point in the liquid. The position of the interface at any time t from the initial end of the boat is simply Rt. If we call

$$x_1=x, \quad x_2=t$$

and

$$u_1=x_b, \quad u_2=t.$$

The relation between coordinate systems is

$$x_1=u_1-Ru_2, \quad x_2=u_2 \tag{3}$$

and (2) becomes (4)

$$\frac{\partial c}{\partial u_2}=D\frac{\partial^2 c}{\partial u_1^2}. \tag{4}$$

[1] R. N. Hall, Phys. Rev. **88**, 139 (1952). Dr. Hall suggests that k may be velocity dependent because of diffusion within the solid, and notes a velocity dependence of k on crystallographic orientation.
[2] R. N. Hall, J. Phys. Chem. **57**, 836 (1951).
[3] W. G. Pfann, J. Metals **4**, 747 (1952).
[4] Tiller, Jackson, Rutter, and Chalmers, Acta Metallurgica, **1**, 428 (1953).
[5] Burton, Prim, and Slichter, J. Chem. Phys. **21**, 1987 (1953).

Next, from elementary calculus

$$\frac{\partial c}{\partial u_1}=\frac{\partial c}{\partial x_1}\frac{\partial x_1}{\partial u_1}+\frac{\partial c}{\partial x_2}\frac{\partial x_2}{\partial u_1}$$

$$=\frac{\partial c}{\partial x_1}\cdot 1+\frac{\partial c}{\partial t}\cdot 0,$$

$$\frac{\partial c}{\partial u_2}=\frac{\partial c}{\partial x_1}\frac{\partial x_1}{\partial u_2}+\frac{\partial c}{\partial x_2}\frac{\partial x_2}{\partial u_2}$$

$$=\frac{\partial c}{\partial x}(-R)+\frac{\partial c}{\partial t}.$$

Thus, (4) becomes

$$\frac{\partial c}{\partial t}=D\frac{\partial^2 c}{\partial x^2}+R\frac{\partial c}{\partial x} \tag{5}$$

which is the desired differential equation in a coordinate system moving with the interface. This equation together with the boundary conditions, should be sufficient to completely describe the physical situation set forth in the assumptions. The equation presented by Tiller *et al.* as the form of the transient liquid distribution [i.e., Eq. (7) in reference 3] will satisfy the differential equation only in the limit as $k\rightarrow 0$, which places an important bound on the domain of validity of their solution. Equation (5) will be solved later using transformation calculus, for the pertinent boundary conditions, which will now be discussed.

BOUNDARY CONDITIONS

Consider in Fig. 1 an element of width dx at the interface of cross-sectional area A, moving with respect to the boat and fixed with respect to the interface. Then $c(x,t)$ is a slowly changing function of time whereas $c(x_b,t)$ is rapidly changing in time. By definition of the segregation constant, the flux of solute leaving the interface at the left face is in time dt

$$\varphi_{\text{out, left face}}=c(0,t)kRA\,dt$$

while

$$\varphi_{\text{in, right face}}=c(dx,t)RA\,dt+DA\left.\frac{\partial c}{\partial x}\right|_{dx}dt.$$

The change in the number of solute atoms in the element in time dt is

$$\Delta n=\varphi_{\text{in}}-\varphi_{\text{out}}$$

$$=DA\left.\frac{\partial c}{\partial x}\right|_{dx}dt+c(dx,t)RA\,dt-c(0,t)kRA\,dt,$$

$$\frac{\Delta n}{\Delta t}=DA\left\{\left.\frac{\partial c}{\partial x}\right|_0+dx\frac{\partial^2 c}{\partial x^2}\right\}+RA\left\{c(0,t)+dx\frac{\partial c}{\partial x}-c(0,t)k\right\}$$

$$=DA\left\{\frac{\partial c}{\partial x}+dx\frac{\partial^2 c}{\partial x^2}\right\}+RA\left\{(1-k)c(0,t)+\frac{\partial c}{\partial x}dx\right\},$$

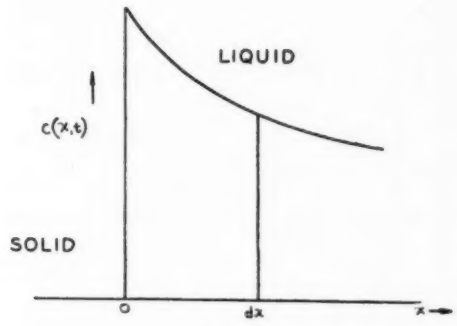


FIG. 1.

now

$$\Delta c=\frac{\Delta n}{\text{volume}}=\frac{\Delta n}{A\,dx}$$

so

$$\frac{\Delta c}{\Delta t}=D\frac{\partial c}{\partial x}\frac{1}{\partial x}+D\frac{\partial^2 c}{\partial x^2}+Rc(0,t)[1-k]\frac{1}{dx}+R\frac{\partial c}{\partial x}$$

but, we know, Eq. (5)

$$\frac{\Delta c}{\Delta t}=D\frac{\partial^2 c}{\partial x^2}+R\frac{\partial c}{\partial x};$$

therefore

$$\frac{1}{dx}\left[D\frac{\partial c}{\partial x}+R(1-k)c(0,t)\right]=0$$

or

$$\left.D\frac{\partial c}{\partial x}\right|_{x=0}=-R(1-k)c(0,t) \tag{6}$$

corresponding to the requirement that the concentration gradient at the interface must be sufficient to retain the rejected solute in the liquid phase.

It should be noticed at this point that the equation proposed in reference 3, [Eq. (7) in reference 3], does not satisfy Eq. (6).

A further useful condition is obtained from the integral representation of the fact that the solute rejected from the solid must be equal to that gained by the melt. Graphically, this requires that the cross-hatched areas in Fig. 2 be equal. Mathematically, conservation of solute can be expressed by

$$R\int_0^t[C_0-kc(0,t)]dt=\int_0^L[c(x,t)-C_0]dx \tag{7}$$

for the first pass when the initial solute distribution is not a function of distance, and where L is the length of the boat for the case of normal freezing or is equal to the length of the molten zone in zone purification. Equation (7) can be generalized for the case of repeated passes by utilizing the previous distribution so that for

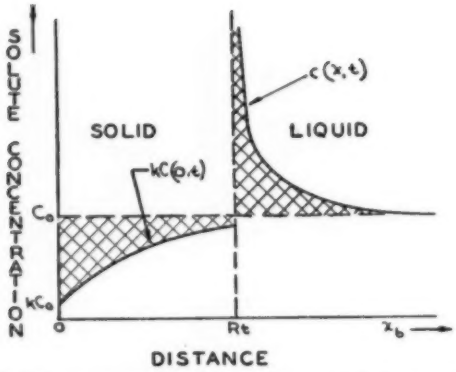


FIG. 2. Solute concentration in liquid and solid phases after interface has propagated a distance Rt through the crystal.

the nth pass

$$kR\int_0^t [c^{n-1}(0,t)-c^n(0,t)]dt = \int_0^L [c^n(x,t)-c^{n-1}(x,t)]dx, \quad (8)$$

where $c^i(x,t)$ represents the solute distribution in the liquid for the ith pass. Equations (5), (6), and (8) are the primary equations to be considered in obtaining the solution to the problem, for the set of assumptions set forth above.

TRANSFORMATION AND SOLUTION

Transformation calculus employing Laplace transforms seems to be particularly suitable for obtaining the solution to the equations because the boundary conditions are utilized in the course of the solution, rather than trying to obtain the most general solution initially and particularizing the result to the special case afterward. The Laplace transform $F(s)$ of a function $f(t)$ is defined by

$$F(s)\equiv\int_0^\infty f(t)e^{-st}dt\equiv L[f(t)]. \quad (9)$$

This shall be utilized to transform Eq. (5), (6), and (8) to functions of x and s. From a table of transforms[6]

$$L\left[\frac{\partial c}{\partial t}\right]=sF(s)-c(x,t=0+)$$
$$=sc(x,s)-c(x,t=0+)$$

and

$$L\left[D\frac{\partial^2 c}{\partial x^2}\right]=DL\left[\frac{\partial^2 c}{\partial x^2}\right]=D\frac{\partial^2}{\partial x^2}L[c(x,t)]$$
$$=D\frac{\partial^2}{\partial x^2}c(x,s)$$
$$=D\frac{d^2}{dx^2}c(x,s),$$

where the last step is allowable since s is not a function of x. Operating in a similar manner on the velocity term wherein the assumption that R is not a function of time is utilized, the differential equation (5) becomes

$$D\frac{d^2}{dx^2}c(x,s)+R\frac{d}{dx}c(x,s)-sc(x,s)=-c(x,t=0+). \quad (10)$$

For the first pass which shall be considered first, the right-hand side of (10) is C_0, so that

$$\frac{d^2c}{dx^2}+\frac{R}{D}\frac{dc}{dx}-\frac{s}{D}c=-\frac{C_0}{D} \quad (11)$$

which is a total rather than partial differential equation having the customary solution

$$C=M\exp(\alpha^+x)+N\exp(\alpha^-x)+\text{P.I.} \quad (12)$$

where α^+ and α^- are the roots of the quadratic, namely

$$\alpha^\pm=-\frac{R}{2D}\pm\left(\frac{R^2}{4D^2}+\frac{s}{D}\right)^{\frac{1}{2}} \quad (13)$$

and P.I. represents the particular integral corresponding to the right-hand side of (11). The P.I. is easily found by assuming a power series solution in x and equating coefficients on both sides which yields for the first pass

$$\text{P.I.}=\frac{C_0}{s}$$

so that

$$c(x,s)=M\exp(\alpha^+x)+N\exp(\alpha^-x)+\frac{C_0}{s} \quad (14)$$

in which M and N are functions of s. The inverse transformation of $c(x,s)$ back to $c(x,t)$ cannot be performed until M and N are determined utilizing the boundary conditions. Since we do not expect our solution to increase without limit for large values of x we set $M=0$.

The evaluation of N can be achieved either with the integral or differential representation of the boundary conditions. The transformation of the differential boundary condition yields

$$\left.\frac{\partial c(x,s)}{\partial x}\right|_{x=0}=-\beta c(0,s) \quad (15)$$

where

$$\beta\equiv[1-k]\frac{R}{D}. \quad (16)$$

The integral boundary condition (7) is transformed with the help of a table of transforms term by term yielding for the case of $L=\infty$ and demanding $c(\infty,t)=C_0$,

$$\frac{RC_0}{s^2}-\frac{kR}{s}c(0,s)=\int_0^L\left[c(x,s)-\frac{C_0}{s}\right]dx, \quad L\to\infty \quad (16a)$$

[6] S. Goldman, *Transformation Calculus and Electrical Transients* (Prentice-Hall, Inc., New York, 1949).

substituting (14) into the above

$$\frac{RC_0}{s^2}-\frac{kR}{s}\left[N+\frac{C_0}{s}\right]=-\frac{1}{\alpha^-}N$$

solving for N

$$N=-\frac{R(1-k)C_0\alpha^-}{s(s-\alpha^-kR)}$$

so that

$$c(x,s)=\frac{C_0}{s}-\frac{R(1-k)C_0\alpha^-}{s(s-\alpha^-kR)}\exp(\alpha^-x). \qquad (17)$$

Equation (17) must now be transformed back to $c(x,t)$ via the relation

$$f(t)=\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}F(s)e^{st}ds. \qquad (18)$$

The contour of the complex integral generally being taken parallel and to the right of the imaginary axis and then completed at ∞ in the negative half-plane. The integral can then be evaluated with the help of the method of residues.

The second term in (17) has a simple pole at the origin, a branch point at $s=-R^2/4D$, and a singularity at $s=-k\beta R$. The latter singularity depends on k in an interesting manner; for $0<k<\frac{1}{2}$ and $\frac{1}{2}<k<1$ the singularity is located between the branch point and the origin; for $k=0$ or $k=1$ it is located at the origin, and for $k=\frac{1}{2}$, it is located at the branch point. The following evaluation of (17) will assume that the singularity is located between the branch point and the origin thus excluding the values $k=0$, $\frac{1}{2}$, and 1 from the applicability of the results.

Before the integration contour in the complex plane is specified, the nature of the singularity at $s=-k\beta R$ must be examined. The functional behavior of the second term in the denominator of the second term in Eq. (17), namely $f(s)=s-\alpha^-kR$, in the vicinity of the singularity determines the order of the pole at that point. If we define $\bar{s}=-k(1-k)R^2/D$, then

$$f(s)=(s-\bar{s})+\frac{kR^2}{2D}+\frac{kR}{\sqrt{D}}\left[(s-\bar{s})+\frac{R^2}{4D}+\bar{s}\right]^{\frac{1}{2}}+\bar{s}$$

letting $\Delta s=s-\bar{s}$

$$\varphi(\Delta s)=f(s)$$

$$=\Delta s+\frac{kR}{\sqrt{D}}\left(\Delta s+\frac{R^2}{4D}+\bar{s}\right)^{\frac{1}{2}}+\frac{kR^2}{2D}+\bar{s}$$

now, expand $\varphi(\Delta s)$ in a Taylor's series

$$\varphi(\Delta s)=\varphi(0)+\frac{\varphi'(0)\Delta s}{1!}+\frac{\varphi''(0)(\Delta s)^2}{2!}+\cdots,$$

where

$$\varphi(0)=\frac{kR^2}{2D}-k(1-k)\frac{R^2}{D}+\frac{kR}{\sqrt{D}}\left(s+\frac{R^2}{4D}\right)^{\frac{1}{2}},$$

$$\varphi(0)=-\frac{kR^2}{2D}(1-2k)+\frac{kR^2}{2D}(1-4k+4k^2)^{\frac{1}{2}},$$

$$=-\frac{kR^2}{2D}(1-2k)+\frac{kR^2}{2D}\{\pm(1-2k)\}$$

since $\varphi(0)=f(s)=0$, the $+$ sign must be chosen to obtain consistency. Next,

$$\varphi'(\Delta s)=1+\frac{kR}{2\sqrt{D}}\frac{1}{\left[\frac{R^2}{4D}-k(1-k)\frac{R^2}{D}\right]^{\frac{1}{2}}}$$

$$=1+\frac{kR}{2\sqrt{D}}\frac{1}{\frac{R}{2\sqrt{D}}[(1-2k)^2]^{\frac{1}{2}}}$$

$$=1+\frac{k}{1-2k}=\frac{1-k}{1-2k} \quad \text{if take } + \text{ root}$$

$$=1-\frac{k}{1-2k}=\frac{1-3k}{1-2k} \quad \text{if take } - \text{ root.}$$

The value for the positive root seems to be more in accord with the previously discussed behavior of the singularity and will be adopted. Hence

$$\varphi(\Delta s)=0+\frac{1-k}{1-2k}\Delta s+\text{higher-order terms.} \qquad (19)$$

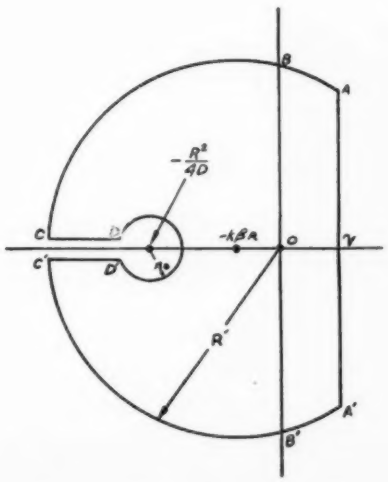


FIG. 3. Contour employed for evaluation of the inversion integral.

Since Δs is small in the vicinity of the singularity, the contribution of higher-order terms in Δs is yet smaller, so that the singularity can be considered a simple pole, and its contribution to the integral is the residue at that point.

With the nature of the singular points established, the contour can now be specified. Extend a branch cut from the branch point to negative infinity along the axis of reals. The contour is then deformed as indicated in Fig. 3 so as not to cross the branch cut. There will be a contribution to the integral from the paths on opposite sides of the branch cut since the function is not continuous across the cut.

Now, by the Cauchy-Goursat theorem and the residue theorem,

$$f(t)\equiv\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}F(s)e^{st}ds$$
$$=-\frac{1}{2\pi i}\{I_{AC}+I_{CD}+I_{DD'}+I_{D'C'}+I_{C'A'}\} \qquad (20)$$
$$+\text{residues at poles}$$

where, I_{AC} signifies the integral taken in the counterclockwise direction from A to C, etc. As we let $R'\to\infty$, $\gamma\to 0$, $I_{AC}=I_{C'A'}\to 0$, because of the negative exponential. For the contribution around the branch point consider the path to be along a small circle of radius r_0, so that $s=r_0e^{i\theta}$ and $[s+(R^2/4D)]^{\frac{1}{2}}=\sqrt{r_0}e^{i(\theta/2)}$. Substituting into Eq. (17),

$$I_{DD'}=\int_{(\pi-\epsilon)}^{-(\pi-\epsilon)}\frac{\left(-\frac{R}{2D}-\frac{1}{\sqrt{D}}\sqrt{r_0}e^{i(\theta/2)}\right)\exp(-(R^2/4D)t)\exp(r_0te^{i\theta})\exp(-(R/2D)x)\times\exp\left(-\frac{1}{\sqrt{D}}\sqrt{r_0}e^{i(\theta/2)}x\right)r_0ie^{i\theta}d\theta}{\left(r_0e^{i\theta}-\frac{R^2}{4D}\right)\left(r_0e^{i\theta}-\frac{R^2}{4D}+\frac{kR^2}{2D}+\frac{kR}{\sqrt{D}}\sqrt{r_0}e^{i(\theta/2)}\right)},$$

$$\lim_{r_0\to 0}I_{DD'}=\int_{(\pi-\epsilon)}^{-(\pi-\epsilon)}\frac{\left(-\frac{R}{2D}\right)\exp(-(R^2/4D)t)\exp(-(R/2D)x)\cdot 0\cdot ie^{i\theta}}{-\frac{R^2}{4D}\left(-\frac{R^2}{4D}+\frac{kR^2}{2D}\right)}d\theta.$$
$$=0$$

Yielding no contribution from the integration around the branch point. Next, proceeding in a similar fashion along the branch cuts

$$I_{CD}=\int_{R'}^{(r_0+R^2/4D)}\frac{\left(-\frac{R}{2D}-\frac{i}{\sqrt{D}}\left(r-\frac{R^2}{4D}\right)^{\frac{1}{2}}\right)e^{-tr}e^{-(R/2D)x}\exp\left(-\frac{ix}{\sqrt{D}}\left(r-\frac{R^2}{4D}\right)^{\frac{1}{2}}\right)dr}{-r\left(-r-\left[-\frac{R}{2D}-\frac{i}{\sqrt{D}}\left(r-\frac{R^2}{4D}\right)^{\frac{1}{2}}\right]kR\right)},$$

$$I_{D'C'}=\int_{r_0+R^2/4D}^{R'}\frac{\left[-\frac{R}{2D}+\frac{i}{\sqrt{D}}\left(r-\frac{R^2}{4D}\right)^{\frac{1}{2}}\right]e^{-tr}e^{-(R/2D)x}\exp\left[\frac{ix}{\sqrt{D}}\left(r-\frac{R^2}{4D}\right)^{\frac{1}{2}}\right]dr}{-r\left(-r-\left[-\frac{R}{2D}+\frac{i}{\sqrt{D}}\left(r-\frac{R^2}{4D}\right)^{\frac{1}{2}}\right]kR\right)},$$

$$\lim_{\substack{r_0\to 0\\ R'\to\infty}}(I_{CD}+I_{D'C'})=e^{-(Rx/2D)}\int_{R^2/4D}^{\infty}\frac{e^{-tr}}{r}\left\{\frac{R}{2D}\left[\frac{\exp\left(-\frac{ix}{\sqrt{D}}\left(r-\frac{R^2}{4D}\right)^{\frac{1}{2}}\right)}{-r+\frac{ikR}{\sqrt{D}}\left(r-\frac{R^2}{4D}\right)^{\frac{1}{2}}+\frac{kR^2}{2D}}-\frac{\exp\left(\frac{ix}{\sqrt{D}}\left(r-\frac{R^2}{4D}\right)^{\frac{1}{2}}\right)}{-r-\frac{ikR}{\sqrt{D}}\left(r-\frac{R^2}{4D}\right)^{\frac{1}{2}}+\frac{kR^2}{2D}}\right]\right.$$
$$\left.+\frac{i}{\sqrt{D}}\left(r-\frac{R^2}{4D}\right)^{\frac{1}{2}}\left[\frac{\exp\left(-\frac{ix}{\sqrt{D}}\left(r-\frac{R^2}{4D}\right)^{\frac{1}{2}}\right)}{-r+\frac{ikR}{\sqrt{D}}\left(r-\frac{R^2}{4D}\right)^{\frac{1}{2}}+\frac{kR^2}{2D}}+\frac{\exp\left(\frac{ix}{\sqrt{D}}\left(r-\frac{R^2}{4D}\right)^{\frac{1}{2}}\right)}{-r-\frac{ikR}{\sqrt{D}}\left(r-\frac{R^2}{4D}\right)^{\frac{1}{2}}+\frac{kR^2}{2D}}\right]\right\}dr.$$

Even with the substitution $\mu^2=r-(R^2/4D)$, an exact integration of the above form could not be found. To simplify the integrals, assume that k and R are small so that the imaginary terms in the denominator can be neglected. Then

$$I_{CD}+I_{D'C'}=\int_{R^2/4D}^{\infty}\frac{e^{-Rx/2D}e^{-tr}}{r\left(-r+\frac{kR^2}{2D}\right)}\left\{-\frac{iR}{D}\sin\frac{x}{\sqrt{D}\left(r-\frac{R^2}{4D}\right)^{\frac{1}{2}}}+\frac{2i}{\sqrt{D}\left(r-\frac{R^2}{4D}\right)^{\frac{1}{2}}}\cos\frac{x}{\sqrt{D}\left(r-\frac{R^2}{4D}\right)^{\frac{1}{2}}}\right\}dr. \quad (21)$$

This can be written, term by term, as

$$I=\int_a^{\infty}e^{-tr}f(r)S(r)dr,$$

where $S(r)$ represents the trigonometric term, and $f(r)$ the algebraic terms. If we now expand $f(r)$ in a Mac-Laurin's series, then

$$I\simeq\int_a^{\infty}f(a)e^{-tr}S(r)dr+\int_a^{\infty}(r-a)\frac{f'(a)}{1!}e^{-tr}S(r)dr+\cdots$$

hoping that the exponential term drops off sufficiently fast that the contributions are large only for small r, r near a, such that one can expand about that point. For the first term in (21)

$$f(r)=-\frac{iR}{D}e^{-Rx/2D}\left[1\Big/r\left(-r+\frac{kR^2}{2D}\right)\right],$$

$$f(a)=-\frac{iR}{D}e^{-Rx/2D}\left[1\Big/\left(\frac{R^2}{4D}\right)^2(1-2k)\right],$$

$$f'(r)=\frac{iR}{D}e^{-Rx/2D}\left[\left(2r-\frac{kR^2}{2D}\right)\Big/r^2\left(r-\frac{kR^2}{2D}\right)^2\right],$$

$$f'(a)=\frac{iR}{D}e^{-Rx/2D}\left[\frac{R^2}{2D}(1-k)\Big/\left(\frac{R^2}{4D}\right)^4(1-2k)^2\right],\quad\text{etc.}$$

For the zeroth-order term, the contribution is, making the substitution

$$\mu^2=r-\frac{R^2}{4D}\equiv r-a,$$

$$\exp\left(-\frac{R^2t}{4D}\right)\int_0^{\infty}f(a)2\mu\sin\frac{x}{\sqrt{D}}\mu\exp(-t\mu^2)d\mu=2f(a)\exp\left(-\frac{R^2t}{4D}\right)\left\{\frac{x}{\sqrt{D}}\frac{\sqrt{\pi}}{4t^{\frac{1}{2}}}\right\},$$

where the integration is obtained from tables.[7] The higher-order terms and the cosine terms are obtained in the same manner, yielding, up to first order,

$$I_{CD}+I_{D'C'}=\frac{4i}{R}\left(\frac{\pi}{Dt}\right)^{\frac{1}{2}}\frac{4D}{R^2t}\exp\left(-\frac{x^2+2xRt+R^2t^2}{4Dt}\right)\times\left\{x\left[1+\frac{1-k}{1-2k}\frac{4D}{R^2}\frac{6t-(x^2/D)}{4t^2}+\cdots\right]+\frac{2D}{R}\left[1-\frac{x^2}{2Dt}+\frac{1-k}{1-2k}\frac{4D}{R^2}\frac{12t^2-12t(x^2/D)+x^4/D^2}{32t^3}+\cdots\right]\right\}. \quad (22)$$

The contributions at the poles are easily determined by evaluating the residues at these points. The residue at the origin is

$$-\frac{\alpha^-(0)\exp(\alpha^-(0)x)}{\alpha^-(0)kR}=-\frac{1}{kR}e^{-(R/D)x}.$$

At the other pole, $s=\alpha^-kR=-k\beta R$, the residue is, employing Eq. (19)

$$\frac{\alpha^-(-k\beta R)}{kR\alpha^-(-k\beta R)}\exp(\alpha^-(-k\beta R)x)\exp(kRT\alpha^-(-k\beta R))\times\left(\frac{1-2k}{1-k}\right)=\frac{1}{kR}e^{-\beta x}e^{-k\beta Rt}\left(\frac{1-2k}{1-k}\right). \quad (23a)$$

Then, since the inverse transform of C_0/s is just C_0, combining the various contributions, the following form is obtained for the time dependent spatial distribution of solute in the liquid:

$$c(x,t)=C_0\left\{1+\frac{1-k}{k}\left[e^{-(R/D)x}-\frac{1-2k}{1-k}e^{-(1-k)(R/D)x}\times e^{(1-k)(kR/D)Rt}\right]+\frac{2}{\pi}\left(\frac{\pi}{Dt}\right)^{\frac{1}{2}}\frac{4D}{R^2t}\frac{1-k}{1-2k}\times\exp\left(-\frac{(x+Rt)^2}{4Dt}\right)\left[x\left\{1+\frac{1-k}{1-2k}\frac{4D}{R^2}\times\frac{6t-(x^2/D)}{4t^2}+\cdots\right\}+\frac{2D}{R}\left\{1-\frac{x^2}{2Dt}+\frac{1-k}{1-2k}\frac{4D}{R^2}\frac{12t^2-12t(x^2/D)+x^4/D^2}{32t^3}+\cdots\right\}\right]\right\}. \quad (24)$$

DISCUSSION

By inspection of (24) it is seen that for small D the last term $\to 0$; then assuming k small such that

[7] D. Bierens de Haan, *Nouvelles Tables d'Integrales Definies* (G. E. Stechert and Company, New York, 1939), Table 362, No. 1, p. 506.

$1-2k/1-k\sim 1$, Eq. (24) tends to the form

$$c(x,t)=C_0\left\{1+\frac{1-k}{k}\left[e^{-(R/D)x}-e^{-(1-k)(R/D)x}\times e^{-(1-k)(kR/D)Rt}\right]\right\}. \quad (25)$$

To determine the distribution of solute in the solid, set $x=0$, multiply by k, and let $x_b=Rt$ representing distance in the solid measured with respect to the boat from the end initially freezing out, yielding

$$c_s(x_b)\cong kC_0\left\{1+\frac{1-k}{k}\left[1-e^{-(1-k)(kR/D)x_b}\right]\right\} \quad (26)$$

which is similar to the equation obtained by Tiller *et al.*, [Eq. (1) in the foregoing], the prime difference occurring in the exponent, and this difference is sufficient to obtain satisfaction of the boundary condition at the interface, Eq. (6), and of the differential equation, Eq. (5).

By examination of (25) and (26) it is seen that the concentration of solute in the liquid at the interface rises exponentially from its initial value C_0 to the value C_0/k after a long time, while the solid is purified by a factor of $1/k$ initially but progressively less along the ingot, rising at a rate determined by k, R, and D. For effective purification k and R should be small, while D is approximately constant for most liquid metals. Equations and graphs have been presented in the literature expressing the segregation constant as a function of stirring velocities,[8] and growth rates.[5,9] This procedure is based on the definition of an "effective segregation constant" as the ratio of the solute concentration in the solid to the solute concentration in the liquid far from the interface. Care must be taken in interpreting these graphs since such a definition will in practice depend on the length of the boat, turbulence, and where, in practice, the measurements are made on the recrystallized ingot.

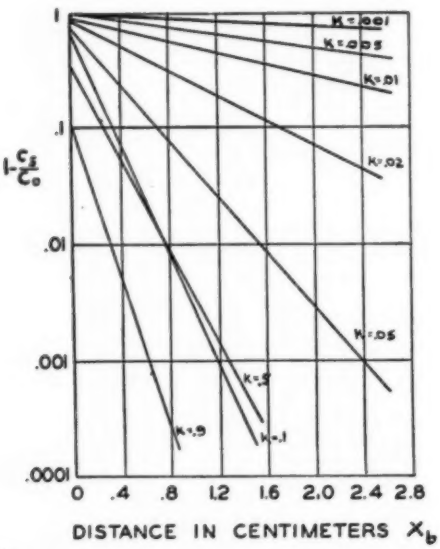


FIG. 4. Plots of the function $[(1-(C_s/C_0))]$ *vs* distance along the recrystallized crystal for various values of the segregation constant. Calculated from Eq. (26) with $R=2\times10^{-3}$ cm/sec and $D=3\times10^{-5}$ cm²/sec.

A useful plot of Eq. (26) can be obtained by plotting $1-(C_s/C_0)$ *vs* distance on semilog paper, as in Fig. 4. For large k the intercept on the $x=0$ line would yield k directly; for small k the slope of the straight line would yield k. Or if k and R are known, the value of the diffusion coefficient, D, could be determined. There is an interesting phenomenon exhibited by the graphs of Fig. 4. The intersection of the $k=0.5$ line with the $k=0.1$ line would imply that there would be greater purification at large distances for the former than the latter value of k. Other than a possibility existing that for the smaller value of k solute is rejected to the melt faster than it can diffuse away, such behavior would seem rather unusual. Perhaps the best procedure is to remember that (26) was derived for small k so that the lines indicated for large k should not be taken as representing the physical situation actually existing.

A quantitative grasp of the extent to which the solute concentration builds up in the liquid ahead of the advancing interface can be obtained from Table I. The solute concentration is approximately normal only 1 millimeter ahead of the interface, whereas the exponential increase of the value near the interface is rather small only after moving about 10 centimeters along the boat, indicating the dubious value of a time independent solution of the diffusion equation.

Table I is also of assistance in understanding the production of p-n junctions by cycling the growth rate of a crystal. Suppose 2 impurities are present in a semiconductor in approximately equal quantities, one of which is a donor with k small, say 0.001, the other an acceptor impurity with k large, $k\sim1$. If the interface advances slowly it will remove the n type donors so that the solid is p type, while the liquid ahead of the interface is strongly n type. If the system is now quickly frozen, i.e., by increasing R, a p-n junction will be created. If the system is then operated at its original velocity, the donor concentration at the new location of the interface will be considerably less because of the sharp concentration drop exhibited in the liquid, so that the segregation process will again produce p-type material thus establishing a p-n-p system which could be extended even further by a similar recycling procedure. Accidental variations in interface velocity caused by mechanical imperfections in the purification apparatus would yield similar results.

If in the inverse transformation of Eq. (17) it is assumed that the singularity at $s=-k\beta R$ is a simple pole and eliminate the Taylor series expansion about that point, and that the contribution adjacent to the branch cut is small, the following form of the solute

[8] W. P. Slichter and B. B. Kolb, Phys. Rev. **90**, 987 (1953).
[9] R. N. Hall, Phys. Rev. **88**, 139 (1952).

distribution in the liquid at a distance x ahead of the interface at a time after starting t is found:

$$c(x,t)=C_0\left\{1+\frac{1-k}{k}\left[e^{-(R/D)x}-e^{-(1-k)(R/D)x}\times e^{-k(1-k)(R/D)Rt}\right]\right\}. \quad (27)$$

The values computed by this relation are tabulated in Table I below the values obtained from the more exact relation, Eq. (24). For the values of the physical constants chosen, which are probably typical to those existing in practice, the numerical differences in the calculated results are of small magnitude. The solute distribution in the recrystallized solid corresponding to this approximate solution is exactly Eq. (26) above.

Turbulent Solution

Burton, Prim, and Slichter[5] suggest a different form for the solution of the basic differential equation, Eq. (5) in the foregoing, by restricting the diffusion process to a distance δ away from the interface, corresponding to that distance where the laminar fluid flow of the boundary layer region becomes turbulent owing to mixing and convection currents in the melt. The solution is facilitated in this case by making the substitutions

$$Y\equiv\delta-x, \quad (28)$$

$$\bar{C}\equiv C(Y,t)-C_{ss}, \quad (29)$$

where $\bar{C}$ is the transient solution, and C_{ss} the steady-state solution of the differential equation, namely, for an infinite boat.

$$C_{ss}(x)=C_0\left\{1+\frac{1-k}{k}\left[e^{-(R/D)x}-e^{-(R/D)\delta}\right]\right\} \quad (30)$$

and

$$C_{ss}(Y)=C_0\left\{1+\frac{1-k}{k}e^{-R\delta/D}\left[e^{RY/D}-1\right]\right\}. \quad (31)$$

With these substitutions, (5) and (6) become

$$D\frac{\partial^2\bar{C}}{\partial Y^2}-R\frac{\partial\bar{C}}{\partial Y}=\frac{\partial\bar{C}}{\partial t} \quad (32)$$

and

$$\left.\frac{\partial\bar{C}}{\partial Y}\right|_{Y=\delta}=\beta\bar{C}(\delta,t). \quad (33)$$

At this point a boundary condition is specified [reference 5, Eq. (18) and Eq. (24)] which requires that there be no solute initially in the region of laminar flow, which is probably difficult to attain in actuality.

TABLE I. Calculated values of the solute concentration in the liquid ahead of an advancing liquid-solid interface for four locations of the interface. Upper number in box represents exact solution, Eq. (24); lower number in box an approximate solution, Eq. (27). Calculated for $k=0.01$, $R=2\times10^{-4}$ cm/sec and $D=3\times10^{-5}$ cm²/sec.

		$\frac{c(x,t)}{C_0}$ in liquid ahead of interface			
	X_b	$X=0.001$ cm	$X=0.01$ cm	$X=0.1$ cm	$X=1$ cm
Distance of interface from beginning of boat	0.2 cm	14.46	7.46	1.0079	1.0000
		13.89	6.39	0.99991	1.0000
	1 cm	46.25	25.67	1.056	1.0000
		45.62	24.02	1.046	1.0000
	6 cm	91.87	50.82	1.124	1.0000
		91.78	50.82	1.123	1.0000
	15 cm	93.61	51.79	1.126	1.0000
		93.56	51.81	1.126	1.0000

Hence, let us specify instead

$$C(x,0)=C_0 \quad \text{for } 0<x<\delta \quad (34)$$

$$C(\delta,t)=C_0 \quad (35)$$

so that

$$\bar{C}(Y,0)=C_0-C_{ss}(Y)=-C_0\frac{1-k}{k}e^{-R\delta/D}\left[e^{RY/D}-1\right] \quad (36)$$

in which case, we automatically have

$$\bar{C}(0,t)=0. \quad (37)$$

Using solutions of (32) of the form

$$\bar{C}_\mu(Y,t)=B_\mu e^{RY/2D}\sin\mu Y\exp\left(-\frac{R^2t}{4D}\right)\exp(-\mu^2Dt) \quad (38)$$

wherein to satisfy the boundary condition at the interface, (33), the μ's must be roots of the equation

$$\mu\delta=(\tfrac{1}{2}-k)\frac{R\delta}{D}\tan\mu\delta \quad (39)$$

which has an infinite number of roots, positive and negative. Proceeding as in reference 5, denote the ith positive root by μ_i. Then the arbitrary coefficients B in (38) are determined by the requirement of having to construct an infinite series into an initial solution that will be equal in magnitude and opposite in sign to the

spatial dependence of the steady-state solution.

$$e^{RY/2D}\sum_{i=1}^{\infty} B_i \sin\mu_i Y = C_0 - C_{ss} \tag{40}$$

and using orthogonality,

$$B_i = \frac{\int_0^{\delta} e^{-RY/2D}[C_0 - C_{ss}]\sin\mu_i Y dY}{\int_0^{\delta}\sin^2\mu_i Y dY}. \tag{41}$$

Then carrying out the indicated integrations, we obtain,

defining $\Delta \equiv R\delta/D$,

$$\bar{C}(Y,t) = -4C_0 e^{RY/2D}\frac{1-k}{k}e^{-R\delta/D}$$
$$\sum_{i=1}^{\infty}\frac{\left\{\frac{R\delta}{2D}\sin\mu_i\delta\cosh\frac{R\delta}{2D} - \mu_i\delta\cos\mu_i\delta\sinh\frac{R\delta}{2D}\right\}}{\left(1-\frac{\sin 2\mu_i\delta}{2\mu_i\delta}\right)[(\Delta/2)^2+(\mu_i\delta)^2]}$$
$$\times\sin\mu_i Y\exp(-R^2t/4D)\exp(-\mu_i^2 Dt). \tag{42}$$

From which, by setting $Y=\delta$ and $x_b = Rt$ the solute distribution inthe solid as a function of distance along the boat, is found to be

$$C_s(x_b) = kC_0\left\{1+\frac{1-k}{k}e^{-R\delta/2D}\left[2\sinh\frac{R\delta}{2D}\right.\right.$$
$$\left.\left. -4\delta e^{-Rx_b/4D}\sum_{i=1}^{\infty}\frac{\left\{\frac{R\delta}{2D}\sin\mu_i\delta\cosh\frac{R\delta}{2D} - \mu_i\delta\cos\mu_i\delta\sinh\frac{R\delta}{2D}\right\}\sin\mu_i\delta e^{-\mu_i^2 Dx_b/R}}{\left(1-\frac{\sin 2\mu_i\delta}{2\mu_i\delta}\right)[(\Delta/2)^2+(\mu_i\delta)^2]}\right]\right\}. \tag{43}$$

Numerical calculations with (43) tend to be somewhat tedious so that no numerical comparison of (43) with (26) will be presented. However, the primary difference between the two is seen to be in the difference in the exponentials of distance. The solute distribution given by (43) varies as $\exp(-Rx_b/4D)$ while in (26) it varies approximately as $\exp(-kRx_b/4D)$ which means the latter would yield higher purity over a length of the crystal $(1/k)$ longer than that Eq. (43) would predict. The exponential term contained inside the summation sign of Eq. (43) is small in comparison with the one outside, at least for small μ_i, and so is neglected in the foregoing discussion.

The preceding analysis utilizing Laplace transforms can be applied to consider the results of passing several segregating interfaces through a crystal, the effect of finite boat length, behavior under a velocity variation, and boundary layer effects due to turbulence in the melt. These questions arise in the purification technique known as zone purification and in the growth of crystals by the Czochralski method. The analytical consideration of these problems is feasible but somewhat involved; it is hoped that they can be considered in more detail in a subsequent paper, together with pertinent experimental data.

The author would like to express his thanks to W. C. Davidon of the Nuclear Instrument and Chemical Corporation and to Hans Ekstein of Armour Research Foundation for their friendly assistance with the analysis.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 9 SEPTEMBER, 1954

Matrix Analysis of Linear Time-Varying Circuits

LOUIS A. PIPES
University of California and U. S. Naval Ordnance Test Station, Inyokern, China Lake, California
(Received December 7, 1953)

This paper presents a general mathematical technique for studying the response and stability of linear time-varying circuits that contain parameters whose magnitudes vary in a periodic manner with the time. The method presented reduces the solution of the circuit differential equation to that of computing powers of matrices. For purposes of illustration, the method is used to determine the response of an electric circuit consisting of a constant resistance and inductance in series with a periodically varying capacitance. The cases of square-wave and saw-tooth-wave variations of the variable parameter are considered in detail.

1. INTRODUCTION

THE general theory of electric circuits whose component parameters are linear and constant has been extensively developed and is well understood. In many technical problems, however, it is desired to determine the effect of a circuit element whose value varies with the time alone. Such circuits are called linear, time-varying circuits, to distinguish them from nonlinear circuits that contain parameters whose values vary with the current that passes through them or the potential that exists across them. Examples of linear, time-varying circuits of practical importance are the microphone transmitter, which contains a variable resistance, whose value is varied by some source of energy outside the circuit; the condenser microphone, which contains a variable capacitor; the induction generator, in which the mutual inductance between the primary and the secondary is varied.

In recent years, attention has been paid to the behavior of circuits whose parameters vary *periodically* with the time. These circuits are of importance in the theory of frequency modulation,[1,2] superregeneration,[3] and the parametric excitation of oscillations.[4–8] Analogous systems in mechanical and acoustical engineering in which the compliance or inertia parameters vary periodically with the time, lead to the same mathematical formulation.[9,10] The mathematical formulation of the behavior of the circuits enumerated above involves the determination of the solution of the Mathieu-Hill differential equation,

$$\frac{d^2x}{dt^2}+xF(t)=0, \tag{1.1}$$

in which $F(t)$ is a given periodic function of t and suitable initial conditions are prescribed. The theoretical foundations of the usual studies encountered in the literature are based on the classical Floquet theory of Hill's equation.[11] In the present discussion, an extension of a different method given[12] will be presented and illustrated by its application to a circuit involving capacitative parametric excitation.

2. THE GENERAL SERIES CIRCUIT WITH VARIABLE CAPACITANCE

In order to illustrate the general mathematical technique to be presented, consider the circuit illustrated by Fig. 1.

This circuit consists of a constant resistance and inductance parameters R and L in series with a time-varying capacitance whose elastance is $S(t)$. It is desired to determine the behavior of the circulating charge $q(t)$ when the type of periodic variation of the elastance $S(t)$ and the initial charge separation and current of the circuit are given at $t=0$. Since there is a possibility of energy being furnished to the circuit by the agent that is causing the time variation of the capacitor, this circuit is said to be parametrically excited. An analysis of the response of this circuit by the classical Floquet theory of Mathieu's equation will be found in the paper by Barrow, Smith, and Baumann.[6]

If Kirchhoff's second law is applied to the circuit of Fig. 1, the following differential equation for the mesh charge is obtained,

$$L\ddot{q}+R\dot{q}+S(t)q=0. \tag{2.1}$$

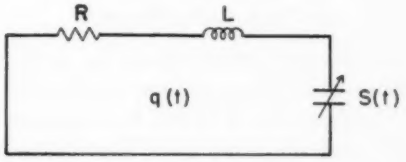


FIG. 1. General series circuit.

[1] J. R. Carson, Proc. Inst. Radio Engrs. **10**, 62 (1922).
[2] B. van der Pol, Proc. Inst. Radio Engrs. **18**, 1194–1205 (1930).
[3] P. Bura and D. M. Tombs, Wireless Engr. 95–100 (April, 1952); 120–126 (May, 1952).
[4] M. Brillouin, Eclairage Elect. **XI**, 49–59 (April, 1897).
[5] A. Erdelyi, Ann. Physik **23**, 21 (1935).
[6] Barrow, Smith, and Baumann, J. Franklin Inst. **221**, 403–529 (1936).
[7] N. Minorsky, J. Franklin Inst. **240**, No. 1, 25–46 (July, 1945).
[8] E. I. Hawthorne, Proc. Inst. Radio Engrs. **39**, 78–81 (1951).
[9] S. Timoshenko, *Vibration Problems in Engineering* (D. Van Nostrand Company, Inc., New York, 1937), pp. 167–181.
[10] A. Stephenson, Proc. Manchester Phil. Soc. **52**, No. 8 (1908).
[11] N. W. McLachlan, *Theory and Application of Mathieu Functions* (Oxford University Press, London, 1946).
[12] L. A. Pipes, J. Franklin Inst. **224**, 767–777 (1937).

It will be assumed that the elastance $S(t)$ is a periodic function of the time of fundamental period T, so that

$$S(t+T)=S(t). \tag{2.2}$$

The initial conditions of the circuit at $t=0$, will be assumed to be given by

$$i(0)=\dot{q}(0)=i_0, \quad q(0)=q_0. \tag{2.3}$$

It is required to solve Eq. (2.1) subject to the initial conditions (2.3) when the functional variation $S(t)$ of the elastance parameter is given explicitly. Before discussing special cases, it is of interest to consider a general formulation of the solution.

Equation (2.1) may be written in the following form,

$$\ddot{q}+2b\dot{q}+S(t)\frac{q}{L}=0, \tag{2.4}$$

where

$$b=R/2L. \tag{2.5}$$

To simplify (2.4), it is convenient to introduce the new variable $x(t)$ by the following substitution,

$$q(t)=e^{-bt}x(t). \tag{2.6}$$

In terms of $x(t)$, Eq. (2.4) is transformed into

$$\frac{d^2x}{dt^2}+F(t)x=0, \tag{2.7}$$

where the function $F(t)$ is given by

$$F(t)=\frac{S}{L}(t)-b^2. \tag{2.8}$$

The initial condition (2.3) may be expressed as initial conditions on $x(t)$, by the following equations,

$$x(0)=q_0=x_1(0), \quad \dot{x}(0)=(i_0+bq_0)=x_2(0). \tag{2.9}$$

Since $S(t)$ is a periodic function of fundamental period T the function $F(t)$ is also a periodic function of fundamental period T. The solution of the problem thus resolves itself into a solution of an equation of the Hill type (2.7) subject to the initial conditions (2.9). A general method for the solution of (2.7) will now be discussed.

3. THE CASE OF GENERAL PERIODIC VARIATION OF THE CAPACITANCE

Before considering explicit results for particular types of capacitance variation, a general method for the solution of (2.7) subject to (2.9) will be presented.

Let $u_1(t)$ and $u_2(t)$ be two linearly independent solutions of the Eq. (3.1) in the fundamental interval $0\leq t\leq T$.

$$\frac{d^2x}{dt^2}+F(t)x=0. \tag{3.1}$$

The values of $x(t)=x_1(t)$, and its first derivative, $\dot{x}(t)=x_2(t)$, may be expressed in the following form,

$$\begin{aligned} x_1(t)&=K_1u_1(t)+K_2u_2(t) \\ x_2(t)&=K_1\dot{u}_1(t)+K_2\dot{u}_2(t), \end{aligned} \quad \text{for } 0\leq t\leq T \tag{3.2}$$

where K_1 and K_2 are arbitrary constants.

It is convenient to write the two Eqs. (3.2) in the form of the following single matrix equation,

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}=\begin{bmatrix} u_1(t) & u_2(t) \\ \dot{u}_1(t) & \dot{u}_2(t) \end{bmatrix}\cdot\begin{bmatrix} K_1 \\ K_2 \end{bmatrix}. \tag{3.3}$$

To simplify the writing, the following notation will be introduced

$$\begin{aligned} [x(t)]&=\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, \quad [K]=\begin{bmatrix} K_1 \\ K_2 \end{bmatrix} \\ [u(t)]&=\begin{bmatrix} u_1(t) & u_2(t) \\ \dot{u}_1(t) & \dot{u}_2(t) \end{bmatrix}=\begin{bmatrix} u_{11}(t) & u_{12}(t) \\ u_{21}(t) & u_{22}(t) \end{bmatrix}. \end{aligned} \tag{3.4}$$

In terms of this notation, (3.3) may be expressed in the following concise form,

$$[x(t)]=[u(t)][K]. \tag{3.5}$$

The Wronskian W_0,[13] of the two solutions $u_1(t)$ and $u_2(t)$ is a *constant* in the fundamental interval $0\leq t\leq T$ and is given by the following determinant,

$$W=\det[u(t)]=u_1(t)\dot{u}_2(t)-\dot{u}_1(t)u_2(t). \tag{3.6}$$

Since the two solutions $u_1(t)$ and $u_2(t)$ are linearly independent, $W_0\neq 0$. The matrix $[u(t)]$ is therefore nonsingular and has the following inverse,

$$[u(t)]^{-1}=\frac{1}{W_0}\begin{bmatrix} u_{22}(t) & -u_{12}(t) \\ -u_{21}(t) & u_{11}(t) \end{bmatrix}. \tag{3.7}$$

For $t=0$, (3.5) becomes

$$[x(0)]=[u(0)][K]. \tag{3.8}$$

If this is multiplied by $[u(0)]^{-1}$, the result may be written as follows,

$$[K]=[u(0)]^{-1}[x(0)]. \tag{3.9}$$

This determines the column of arbitrary constants in terms of the initial conditions. If (3.9) is now substituted into (3.5), the result is

$$[x(t)]=[u(t)][u(0)]^{-1}[x(0)]. \tag{3.10}$$

At the end of the fundamental interval $t=T$ and the solution is given in terms of the initial conditions by the equation

$$[x(T)]=[u(T)][u(0)]^{-1}[x(0)]. \tag{3.11}$$

[13] L. R. Ford, *Differential Equations* (McGraw-Hill Book Company, Inc., New York, 1933), pp. 66–69.

Cascading the Solution by the Use of Matrices

If the following change in variable

$$\tau=t-nT, \quad \text{where} \quad 0\leqq\tau\leqq T, \quad n=0, 1, 2, 3, \cdots \tag{3.12}$$

is introduced into the Hill Eq. (3.1), the equation is transformed into

$$\frac{d^2x}{d\tau^2}+F(\tau)x=0, \tag{3.13}$$

since, because of the periodicity of $F(t)$, we have

$$F(\tau+nT)=F(\tau). \tag{3.14}$$

This shows that Hill's equation is *invariant* to the change of variable (3.12). It is thus apparent that a solution of the form (3.11) may be obtained in each interval $0<\tau<T$. The *final* values of x_1 and x_2 at the end of one interval of the variation of $F(t)$ are *initial* values of x_1 and x_2 in the following interval.

If the following notation is introduced

$$[M]=[u(T)][u(0)]^{-1}=\begin{bmatrix}A & B\\ C & D\end{bmatrix}. \tag{3.15}$$

Equation (3.11) may be written in the simple form

$$[x(T)]=[M][x(0)]. \tag{3.16}$$

This gives the values of x_1 and x_2 at the end of one period of the variation of $F(t)$ in terms of their values at the beginning of the period. At the end of the second period of the variation of $F(t)$, we have

$$[x(2T)]=[M][x(T)]=[M]^2[x(0)]. \tag{3.17}$$

It is easily seen that at the end of n periods, we have

$$[x(nT)]=[M]^n[x(0)]. \tag{3.18}$$

The solution within the $(n+1)$th interval is given by

$$[x(nT+\tau)]=[u(\tau)][u(0)]^{-1}[x(nT)], \quad \text{where} \quad 0\leqq\tau\leqq T. \tag{3.19}$$

Equation (3.19) is the solution of Hill's equation at any time $t>0$ in terms of the *initial* conditions and two linearly independent solutions in the fundamental interval $0\leqq t\leqq T$. It is seen that this form of solution requires that positive integral powers of the matrix $[M]$ given by (3.15) be computed. A simple method for effecting this computation will now be given.

4. THE COMPUTATION OF INTEGRAL POWERS OF THE MATRIX $[M]$

An essential part of the solution of Hill's equation by the method under consideration involves the computation of positive integral powers of the second-order square matrix

$$[M]=[u(T)][u(0)]^{-1}=\begin{bmatrix}A & B\\ C & D\end{bmatrix}. \tag{4.1}$$

These powers may be most simply computed by the use of Sylvester's theorem.[14] This important theorem of matrix algebra states that if the characteristic or latent roots of the matrix $[M]$ are distinct, and $P([M])$ is any polynomial of $[M]$, then

$$P([M])=P(r_1)[H_0(r_1)]+P(r_2)[H_0(r_2)], \tag{4.2}$$

where

$$[H_0(r_1)]=[F(r_i)]\Big/\frac{(d\Delta)}{dr}_{r=r_i}. \tag{4.3}$$

r_1 and r_2 are the latent roots of $[M]$, $[F(r)]$ is the *adjoint* of the characteristic matrix of $[M]$, and $\Delta(r)$ is the characteristic function of the matrix $[M]$.

Since the Wronskian W_0 of the solution of Hill's equation as given by (3.6) is a constant throughout the interval $0\leqq t\leqq T$, it follows from (3.6) and (3.15) that the determinant of $[M]$ is equal to unity, that is,

$$\det[M](AD-BC)=1. \tag{4.4}$$

As a consequence of this, the characteristic function of the matrix $[M]$ has the following form,

$$\Delta(r)=r^2-r(A+D)+1. \tag{4.5}$$

In order to apply Sylvester's theorem to the computation of $[M]^n$, the polynomial $P([M])$ is taken to be

$$P([M])=[M]^n, \quad n=0, 1, 2, 3, \cdots. \tag{4.6}$$

This gives the required form for powers of $[M]$ in the cases for which $(A+D)\neq 2$. If $(A+D)=2$, both latent roots become equal to unity, and if $(A+D)=-2$ both latent roots become equal to -1. The proper form for $[M]^n$ in these cases may be obtained by the use of a confluent form of Sylvester's theorem.[14]

TABLE I. Positive integral powers of $[M]=\begin{bmatrix}A & B\\ C & D\end{bmatrix}$.

1	$(A+D)\neq\pm 2$ $\cosh(a)=\frac{A+D}{2}$ $r_1=e^a$ $r_2=e^{-a}$	$[M]^n=\frac{1}{S_1}\begin{bmatrix}(S_{n+1}-DS_n) & (BS_n)\\ (CS_n) & (S_{n+1}-AS_n)\end{bmatrix}$ $S_n=\sinh(an), \quad n=0, 1, 2, 3, \cdots$
2	$A=D\neq\pm 1$ $\cosh(a)=A$	$[M]^n=\begin{bmatrix}\cosh(an) & Z_0\sinh(an)\\ \frac{\sinh(an)}{Z_0} & \cosh(an)\end{bmatrix}$ $Z_0=\left(\frac{B}{C}\right)^{\frac{1}{2}}, \quad n=0, 1, 2, 3, \cdots$
3	$(A+D)=+2$ $r_1=+1$ $r_2=+1$	$[M]^n=\begin{bmatrix}n(A-1)+1 & nB\\ nC & n(D-1)+1\end{bmatrix}$ $n=0, 1, 2, 3, \cdots$
4	$(A+D)=-2$ $r_1=r_2=-1$	$[M]^n=nE_{n-1}\begin{bmatrix}(A+1) & B\\ C & (D+1)\end{bmatrix}+E_nI$ $E_n=\exp(jn\pi), \quad I=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$ $n=0, 1, 2, 3, \cdots$

[14] Frazer, Duncan, and Collar, *Elementary Matrices* (Cambridge University Press, Cambridge, 1938), pp. 64–85.

The various possible forms for $[M]^n$ are given in Table I. These expressions for powers of $[M]$ are very useful in the theory of four-terminal networks.[15]

5. SQUARE-WAVE VARIATION OF THE ELASTANCE PARAMETER

The general method discussed above will now be illustrated by applying it to the case in which the elastance parameter $S(t)$ of the circuit of Fig. 1 varies in the manner illustrated by Fig. 2.

In this case, Eq. (2.7) reduces to

$$\begin{aligned} &\frac{d^2x}{dt^2}+F_1x=0, \quad 0\lesseqgtr t\lesseqgtr T/2 \\ &\frac{d^2x}{dt^2}+F_2x=0, \quad T/2\lesseqgtr t\lesseqgtr T, \end{aligned} \tag{5.1}$$

where

$$F_1=\frac{S_1}{L}-b^2, \quad F_2=\frac{S_2}{L}-b^2. \tag{5.2}$$

If the following notation is introduced,

$$w_1=\sqrt{F_1} \quad \text{and} \quad w_2=\sqrt{F_2}. \tag{5.3}$$

The solution of (5.1) may be written in the following matrix form,

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}=\begin{bmatrix} \cos(w_1t) & \dfrac{\sin}{w_1}(w_1t) \\ -w_1\sin(w_1t) & \cos(w_1t) \end{bmatrix}\cdot\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix}, \tag{5.4}$$

where $x_1(0)$ and $x_2(0)$ are the initial values of x, and $\dot{x}$ at $t=0$ given by

$$x_1(0)=q_0, \quad x_2(0)=(i_0+bq_0). \tag{5.5}$$

q_0 is the initial charge and i_0 the initial current of the circuit at $t=0$.

The column on the left of (5.4) gives the value of x and $\dot{x}$ at $t\lesseqgtr T/2$. The values of x and $\dot{x}$ at $t=T/2$ are given by

$$\begin{bmatrix} x_1(T/2) \\ x_2(T/2) \end{bmatrix}=\begin{bmatrix} \cos(\theta_1) & \dfrac{\sin}{w_1}(\theta_1) \\ -w_1\sin(\theta_1) & \cos(\theta_1) \end{bmatrix}\cdot\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix}, \tag{5.6}$$

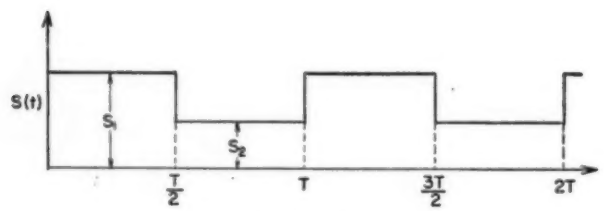


FIG. 2. Rectangular wave.

[15] L. A. Pipes, Phil. Mag. Series 7, **30**, 370–395 (November, 1940).

where

$$\theta_1=\frac{Tw_1}{2}=\frac{T}{2}\sqrt{F_1}. \tag{5.7}$$

Since the values of x and $\dot{x}$ at $t=T/2$ are the *initial* values of x and $\dot{x}$ in the next interval $T/2\lesseqgtr t\lesseqgtr T$, we have for $t=T$

$$\begin{bmatrix} x_1(T) \\ x_2(T) \end{bmatrix}=\begin{bmatrix} \cos(\theta_2) & \dfrac{\sin}{w_2}(\theta_2) \\ -w_2\sin(\theta_2) & \cos(\theta_2) \end{bmatrix} \cdot\begin{bmatrix} \cos(\theta_1) & \dfrac{\sin}{w_1}(\theta_1) \\ -w_1\sin(\theta_1) & \cos(\theta_1) \end{bmatrix}\cdot\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix}, \tag{5.8}$$

where

$$\theta_2=\frac{Tw_2}{2}=\frac{T}{2}\sqrt{F_2}. \tag{5.9}$$

If the indicated matrix multiplication in (5.8) is performed, the result may be written in the following form,

$$\begin{bmatrix} x_1(T) \\ x_2(T) \end{bmatrix}=\begin{bmatrix} A & B \\ C & D \end{bmatrix}\cdot\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix}=[M][x(0)]. \tag{5.10}$$

In this case, the matrix $[M]$ has the form

$$[M]=\begin{bmatrix} A & B \\ C & D \end{bmatrix}, \tag{5.11}$$

where

$$\begin{aligned} A&=\cos(\theta_1)\cos(\theta_2)-\sin(\theta_1)\sin(\theta_2)w_1/w_2 \\ B&=\cos(\theta_2)\sin(\theta_1)/w_1-\cos(\theta_1)\sin(\theta_2)/w_2 \\ C&=-[w_2\cos(\theta_1)\sin(\theta_2)+w_1\cos(\theta_2)\sin(\theta_1)] \\ D&=\cos(\theta_1)\cos(\theta_2)-\sin(\theta_1)\sin(\theta_2)w_2/w_1. \end{aligned} \tag{5.12}$$

The values of x and $\dot{x}$ after the end of n complete periods of the square wave of Fig. 2 are given by

$$[x(nT)]=[M]^n[x(0)]. \tag{5.13}$$

If numerical values are given, the elements A, B, C, D, may be computed by (5.12), and the powers of the matrix $[M]$ may be computed by the formulas of Table I.

At the end of n cycles of the periodic variation of the elastance, the charge and current of the circuit are given by the following equations,

$$\begin{aligned} &q(nT)=e^{-bnT}x_1(nT) \\ &i(nT)=e^{-bnT}[x_2(nT)-bx_1(nT)]. \end{aligned} \qquad n=0, 1, 2, 3, 4, \cdots \tag{5.14}$$

6. THE STABILITY OF THE SOLUTION

One of the most striking properties of time-varying systems is the possibility of exchange of energy between the varying elements and the agents causing their variation. This exchange of energy depends critically on the relative magnitudes of the circuit parameters involved. Experimental investigations[6] have demonstrated that it is possible to generate electrical power in excess of one kilowatt by the proper variation of capacitance.

Hill's equation (1.1) possesses solutions that increase indefinitely with t, and it is interesting to note that Mandelstam and Papalexi[16] developed a circuit containing a periodically time-varying inductance with which it was possible to build up the current in the circuit to such large amplitudes that its insulation was punctured. The determination of the proper adjustments of the circuit parameters to insure stable oscillations leads to a consideration of the stability of the solution of Hill's equation. This matter was treated in a fundamental paper by B. van der Pol and Strutt,[17] by the use of the classical Floquet theory of Hill's equation. It will now be discussed by the use of matrices.

The nature of the solution of Hill's Eq. (1.1) for large values of the independent variable t depends on the behavior of (5.13) for large values of n, and hence depends on the form taken by large powers of the matrix

$$[M]=\begin{bmatrix} A & B \\ C & D \end{bmatrix}. \tag{6.1}$$

In the usual case, we have

$$(A+D)\neq\pm 2, \tag{6.2}$$

and the matrix $[M]^n$ is given by No. 1 of Table I. The circuit charge at $t=nT$ is given by (5.14). This may be written in the following form by the use of Table I.

$$q(nT)=\frac{e^{-bnT}}{S_1}[x_1(0)(S_{n+1}-DS_n)+x_2(0)BS_n], \tag{6.3}$$

where $S_n=\sinh(an)$, and $x_1(0)$ and $x_2(0)$ are given in terms of the initial conditions of the circuit by (5.5). As a consequence of (6.3), it can be seen that for large values of n the charge $q(nT)$ will grow indefinitely with n if

$$a>bT. \tag{6.4}$$

The solution will then be *unstable*. If

$$a<bT, \tag{6.5}$$

the solution will eventually be damped out by the resistance and the oscillations of the circuit will be *stable*. By Table I, we have

$$\cosh(a)=\frac{(A+D)}{2}$$
$$=\cos(\theta_1)\cos(\theta_2)$$
$$-\sin(\theta_1)\sin(\theta_2)(\theta_1^2+\theta_2^2)/2\theta_1\theta_2=k. \tag{6.6}$$

We may therefore write

$$a=\cosh^{-1}(k). \tag{6.7}$$

It can therefore be seen that the oscillations of the circuit will be unstable and grow indefinitely in amplitude with the passage of time provided the circuit parameters are such that

$$\cosh^{-1}(k)>bT. \tag{6.8}$$

The cases for which $(A+D)=\pm 2$ lead to the forms for $[M]^n$ given in entries No. 3 and No. 4 of Table I and may be treated similarly.

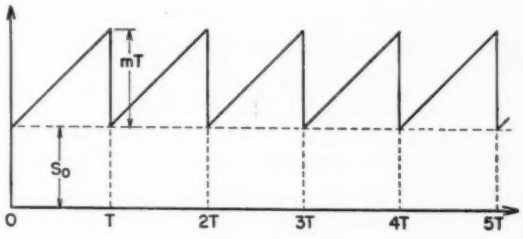


Fig. 3. Sawtooth wave.

7. SAWTOOTH-WAVE VARIATION OF THE ELASTANCE PARAMETER

Another type of variation of the elastance parameter of some practical importance is one in which the parameter undergoes the saw-tooth wave variation illustrated by Fig. 3.

The elastance variation is now given by

$$S(t)=S_{0+mt}, \quad 0\leq t\leq T \tag{7.1}$$

throughout the fundamental interval. In this case, Eq. (2.8) becomes

$$F(t)=\frac{(S_{0+mt})}{L}-b^2, \tag{7.2}$$

or

$$F(t)=At+B, \tag{7.3}$$

where

$$A=\frac{m}{L}, \quad B=\left(\frac{S_0}{L}-b^2\right). \tag{7.4}$$

Throughout the fundamental interval $0\leq t\leq T$ Eq. (2.7) is now given by

$$\frac{d^2x}{dt^2}+(At+B)x=0. \tag{7.5}$$

[16] L. Mandelstam and N. Papalexi, Z. Physik **73**, 223 (1931).
[17] B. van der Pol and M. J. O. Strutt, Phil. Mag. Series 7, **5**, 18 (1928).

This differential equation may be transformed into a simpler one by the following change of the independent variable

$$y=\theta(At+B), \quad \theta=A^{-\frac{1}{3}}. \tag{7.6}$$

This change in variable transforms (7.5) into the equation

$$\frac{d^2x}{dy^2}+yx=0. \tag{7.7}$$

This differential equation is known in the mathematical literature as Stokes' equation and is of considerable importance in mathematical physics because of its occurrence in the theory of diffraction and refraction of waves, as well as in the theory of transmission across potential barriers. Its solution may be expressed in terms of Bessel functions of order one-third in the following form

$$x=(y)^{\frac{1}{2}}\left[C_1J_{\frac{1}{3}}\left(\frac{2y^{\frac{3}{2}}}{3}\right)+C_2Y_{\frac{1}{3}}\left(\frac{2y^{\frac{3}{2}}}{3}\right)\right], \tag{7.8}$$

where C_1 and C_2 are arbitrary constants, and the functions, $J_{\frac{1}{3}}$ and $Y_{\frac{1}{3}}$ are Bessel functions of the first and second kind of order one-third.[18]

For purposes of computation, it is more convenient to express the general solution of Stokes' equation in the following form,

$$x=K_1h_1(y)+K_2h_2(y), \tag{7.9}$$

where h_1 and h_2 are the pair of independent solutions of Stokes' equation which have been tabulated by the Computation Laboratory of Harvard University.[19] K_1 and K_2 are arbitrary constants.

In terms of the functions h_1 and h_2, the general equation (3.2) may be expressed in the following form

$$\begin{aligned} x_1(t)&=K_1h_1(y)+K_2h_2(y),\\ x_2(t)&=K_1\theta Ah_1'(y)+K_2\theta Ah_2'(y), \end{aligned} \tag{7.10}$$

with $y=\theta(At+B)$.

Construct the following matrix,

$$[u(t)]=\begin{bmatrix} h_1(y) & h_2(y)\\ \theta Ah_1'(y) & \theta Ah_2'(y)\end{bmatrix}, \tag{7.11}$$

and let

$$\phi=\theta(AT+B). \tag{7.12}$$

Then

$$[u(T)]=\begin{bmatrix} h_1(\phi) & h_2(\phi)\\ \theta Ah_1'(\phi) & \theta Ah_2'(\phi)\end{bmatrix}. \tag{7.13}$$

We also have

$$[u(0)]=\begin{bmatrix} h_1(\theta B) & h_2(\theta B)\\ \theta Ah_1'(\theta B) & \theta Ah_2'(\theta B)\end{bmatrix}, \tag{7.14}$$

and

$$[u(0)]^{-1}=\frac{1}{\theta AW_0}\begin{bmatrix} \theta Ah_2'(\theta B) & -h_2'(\theta B)\\ -\theta Ah_1'(\theta B) & h_1'(\theta B)\end{bmatrix}. \tag{7.15}$$

W_0 is the Wronskian of the two linearly independent solutions $h_1(y)$ and $h_2(y)$, given by

$$W_0=(h_1h_2'-h_1'h_2)=-1.457j, \quad j=(-1)^{\frac{1}{2}}. \tag{7.16}$$

By (3.15), the matrix $[M]$ is seen to have the following form,

$$\begin{aligned}[M]&=\frac{1}{\theta AW_0}\begin{bmatrix} h_1(\phi) & h_2(\phi)\\ \theta Ah_1'(\phi) & \theta Ah_2'(\phi)\end{bmatrix}\\ &\quad\cdot\begin{bmatrix} \theta Ah_2'(\theta B) & -h_2(\theta B)\\ -\theta Ah_1'(\theta B) & h_1(\theta B)\end{bmatrix}\\ &=\begin{bmatrix} A & B\\ C & D\end{bmatrix}.\end{aligned} \tag{7.17}$$

It may be noticed that as a consequence of the fact that W_0 is a constant, the determinant of $[M]$ is equal to unity as it should be.

Once the numerical values of the elements of the matrix $[M]$ have been determined, the required powers of $[M]$ may be computed by the use of Table I and the circuit charge may be computed by the use of Eq. (6.3). The study of the stability of the solution proceeds in the same manner as that discussed in Sec. VI.

Solution by the Use of Airy Integrals

If the constant part of the elastance S_0 is very large so that the quantity B of (7.4) is large compared to unity, it may be simpler to express the solution of Stokes' Eq. (7.6) in terms of the Airy integrals $A_i(y)$ and $B_i(y)$.[20] These integrals are defined by the two following equations,

$$A_i(y)=\frac{1}{\pi}\int_0^\infty \cos(y\theta+\theta^3/3)d\theta \tag{7.18}$$

$$B_i(y)=\frac{1}{\pi}\int_0^\infty \{\exp(y\theta-\theta^3/3)+\sin(y\theta+\theta^3/3)\}d\theta. \tag{7.19}$$

[18] N. W. McLachlan, *Bessel Functions for Engineers* (Oxford University Press, London, 1934), p. 150.

[19] *Tables of the Modified Hankel Functions of Order One-Third and of Their Derivatives* (Harvard University Press, Cambridge, 1945).

[20] H. Jeffreys and B. S. Jeffreys, *Methods of Mathematical Physics* (Cambridge University Press, Cambridge, 1946), pp. 447–479.

In terms of these integrals, the solution of (7.6) may be expressed in the following form

$$x=k_1A_i(-y)+k_2B_i(-y). \tag{7.20}$$

If $B=(S_0/L-b^2)$ is a large quantity, the absolute value of y remains large in the fundamental interval $0\leqslant t\leqslant T$, and the following asymptotic approximations for the Airy integrals[20] may be used,

$$A_i(-y)=\frac{y^{-\frac{1}{4}}}{(\pi)^{\frac{1}{2}}}\sin\left(\frac{2}{3}y^{\frac{3}{2}}+\pi/4\right) \tag{7.21}$$

and

$$B_i(-y)=\frac{y^{-\frac{1}{4}}}{(\pi)^{\frac{1}{2}}}\cos\left(\frac{2y^{\frac{3}{2}}}{3}+\pi/4\right), \tag{7.22}$$

provided $|y|\gg 1$.

The above asymptotic approximations for the Airy integrals give very good results even for $|y|=1$. If y remains large in the interval $0\leqslant t\leqslant T$, these expressions may be substituted for the functions $h_1(y)$ and $h_2(y)$ in (7.10), and the matrix $[M]$ computed in a manner similar to that indicated above.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 9 SEPTEMBER, 1954

The Statistical Properties of the Output of a Frequency Sensitive Device*

GEORGE R. ARTHUR†
Dunham Laboratory, Yale University, New Haven, Connecticut
(Received June 17, 1953)

Frequently it is desirable to know the statistical properties of the output of certain frequency sensitive devices when some type of random signal is impressed. Since most devices of this type are followed by low pass filters, this problem is one of determining the statistics of a filtered signal.

The problem of finding the probability density of the difference of two dependent random variables has not been solved in any general way. Also, the problem of finding the statistical properties of the output of a linear filter upon which a non-Gaussian random signal is impressed has been considered by only a very few writers.

The first sections of this paper gives a solution of the case of a frequency discriminator with no low pass filter, excited by a narrow band spectrum, and is solved by expressing the output of the device as the difference of two independent random quantities, which allows the use of the characteristic function method.

The later sections contain a solution of the problem when a low pass filter is present at the discriminator output. This is done by obtaining the predominant moments of the density from a series expansion which gives a good approximation to the nature of that density Also, in this regard, the approach of the density to a Gaussian one, as the output filter band is made very narrow, is clearly shown.

1. INTRODUCTORY THEORY

1.1 Frequency Sensitive Devices

CERTAIN frequency-sensitive devices are inherently nonlinear and therefore an investigation of the statistical properties of the output of these devices requires special mathematical techniques. There are numerous methods available for finding the mean value and mean square value of the output of devices of this type.[1–5] These papers also give methods for finding the probability density of the output in the case of no post-detection filter. None of these, however, give any indication of the probability density of the output, when it is filtered, unless the filter band width is extremely narrow.

The frequency-sensitive device to be considered in this paper is a frequency discriminator. This device may be used in some fashion to determine the center frequency of some narrow band input spectrum with certain random characteristics.

1.2 Frequency Discriminator

The frequency discriminator in general has a response characteristic of the form shown in Fig. 1a. In this paper, however, this characteristic will not be used directly, but the discriminator will be represented by the block diagram of Fig. 1b. Here, by spacing the center frequencies and varying the shape of the band pass filters F_1 and F_2 any sort of discriminator characteristic may be obtained. Square law detectors are used here for simplicity.

1.3 Statistical Techniques

Most of the statistical techniques to be used are covered in detail in standard texts on probability and statistics and appropriate references will be cited. The

* This paper is part of a dissertation presented to the Faculty of the Graduate School of Yale University in candidacy for the degree of Doctor of Philosophy and was supported in part by General Precision Laboratory, Pleasantville, New York, under SAR/G.P.L. Sub-contract No. 34.

† Now Sperry Gyroscope Company, Great Neck, New York.

[1] S. O. Rice, Bell System Tech. J. **23**, 282 (1944); **25**, 46 (1945).

[2] D. Middleton, Quart. Appl. Math. **5**, 445 (1948); **7**, 129 (1949); **8**, 59 (1950).

[3] D. Middleton, J. Appl. Phys. **20**, 334 (1949).

[4] F. L. H. M. Stumpers, Proc. Inst. Radio Engrs. **36**, 1080 (1948).

[5] N. M. Blachman, J. Appl. Phys. **20**, 38 (1949).

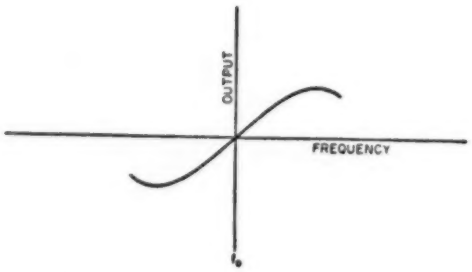


FIG. 1a. General frequency discriminator characteristic.

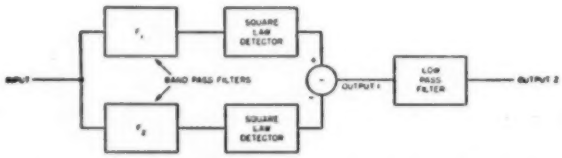


FIG. 1b. Representation of frequency discriminator.

problem of the frequency discriminator is essentially one of finding the probability density of the sum of two dependent random variables. This appears to be a difficult mathematical problem. A general solution to this problem does not seem likely at present, but a solution of some special cases seems reasonable here. Actually, the method to be used is one in which an indication of the statistical properties of the frequency discriminator signal after a filtering device may be found. Then, by letting the filter band width become infinite the direct output of the discriminator, which, in this case, is the difference of two dependent random variables, may be written. The technique to be used will be to set up the problem in such a way that the output may be expressed as the sum of a number of independent random variables whose statistical properties are known. The principle of characteristic functions may be applied to give a solution of the problem.[6] This will be discussed further in the succeeding sections.

2. THE FREQUENCY DISCRIMINATOR-GENERAL

2.1 Representation of Input Signal

In this investigation the type of input signal to be impressed upon the frequency discriminator has a power spectrum which is centered about a frequency f_0 and has a Gaussian shape. The amplitude values are statistically determined, in this case, by a Gaussian probability density. The power spectrum is shown in Fig. 2a.

It is necessary to have some suitable means of representing this input signal. An envelope representation which is reasonable, since the band width of the spectrum is narrow relative to f_0, does not lend itself well to the method to be employed in getting the output in terms of statistically independent quantities. The best approach is to consider a white noise input and shape this noise into the desired spectrum by passing it through suitable filters. Thus, the discriminator will be considered as in Fig. 2b.

2.11 White Noise Input

White noise is characterized by having a completely flat power spectrum. Therefore, it is completely random and uncorrelated. In Fig. 2b the two filters F are chosen so as to have the magnitude of their amplitude transfer equal to the square root of the power spectrum desired. Thus, the actual input to the discriminator is e_1, the signal with the narrow band spectrum. For purposes of analysis, however, it will be most convenient to consider the transfers FF_1 and FF_2 as single filters and consider a single weighting function for each. This approach, of course, has no effect on the validity of the statistical properties found at the output since the discriminator itself is being excited by the type of signal of interest.

Rice[1] has suggested that the noise input may be written in terms of a Fourier series with random coefficients as below:

$$e_1(t)=k\sum_1^K(a_k\cos\omega_k t+b_k\sin\omega_k t), \tag{1}$$

where $\bar{a}_k=\bar{b}_k=0$ and $\langle a_K{}^2=b_K{}^2\rangle_{Av}=W(f)\Delta f$ with the a's and b's Gaussian.

It may be shown that the input signal may be represented by any general orthonormal system with random coefficients. These coefficients are independent and Gaussian on a statistical basis.[7] Represent

$$e_1(t)=\sum_1^\infty \mu_n\phi_n(t) \tag{2}$$

$$\int_{-\infty}^{\infty} dt\phi_n(t)\phi_m(t)=\delta_{mn}. \tag{3}$$

δ_{mn} is the Kronecker delta $\delta_{mn}=\begin{cases}1 & m=n\\ 0 & m\neq n\end{cases}$

$$\mu_n=\int_{-\infty}^{\infty} dte_1(t)\phi_n(t). \tag{4}$$

Using the statistical properties of μ_n one may show, if e_1 is white noise, that

$$\langle\mu_n\rangle_{Av}=0,\quad \langle\mu_n{}^2\rangle_{Av}=W_0/2 \tag{5}$$

with W_0 as power per unit band width. It will be convenient to let $W_0=2$. Thus μ_n is normal with zero mean and standard deviation unity. This will allow easy determination of the probability density of $\mu_n{}^2$ which will be useful later.

2.2 Statistical Properties of Discriminator Output

The voltage symbols to be used in the following refer to those of Fig. 2a. The method to be used here in

[6] H. Cramer, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, 1946).

[7] M. Kac and A. J. F. Siegert, J. Appl. Phys. **18**, 383–397 (1947).

determining the statistical properties of the discriminator output is a modification of one first suggested by Kac and Siegert.[7] Let $Q_1(t)$ be the weighting function of the equivalent filter FF_1 and $Q_2(t)$ that for FF_2.

It will now be convenient to develop the following, without definitely specifying the filter shape. Referring now to Fig. 2a,

$$e_2(t)=\int_{-\infty}^{\infty} d\tau Q_1(t-\tau)e_1(\tau)$$
$$e_3(t)=\int_{-\infty}^{\infty} d\tau Q_2(t-\tau)e_1(\tau). \tag{6}$$

The outputs of the two square-law detectors are

$$e_4(t)=[e_2(t)]^2, \quad e_5(t)=[e_3(t)]^2. \tag{7}$$

The input to the low pass filter is

$$e_6(t)=e_4(t)-e_5(t). \tag{8}$$

If now the weighting function of the low pass filter is denoted as $K(t)$, its output is

$$e_7(t)=\int_{-\infty}^{\infty} d\theta K(t-\theta)[e_4(\theta)-e_5(\theta)]. \tag{9}$$

Substituting in Eq. (9) the values of e_4 and e_5 given by Eqs. (6) and (7) and using the fact that the input is a stationary random process and expressing it as in Eq. (2), one has

$$e_7(t)=\sum_{ij}\mu_i\mu_j\int_{-\infty}^{\infty}\int d\tau_1 d\tau_2\psi(\tau_1,\tau_2)\phi_i(\tau_1)\phi_j(\tau_2), \tag{10}$$

where

$$\psi(\tau_1,\tau_2)=\int_{-\infty}^{\infty} d\theta K(\theta)[Q_1(\tau_1-\theta)Q_1(\tau_2-\theta) - Q_2(\tau_1-\theta)Q_2(\tau_2-\theta)]. \tag{11}$$

Several comments on the foregoing equations are in order at this point. Note that t does not appear on the right side of Eq. (10). This means that $e_7(t)$ is a stationary random process which, of course, is due to the fact that the original input itself was stationary. Also observe that all the integrations of these weighting functions have been carried out from $-\infty$ to ∞. This is mathematically convenient, and all frequency functions and weighting functions will be represented in this fashion. It does not detract from the validity of the statistical results.

Returning now to Eq. (10) consider the integration with respect to τ_2. If the following integral equation

$$\int_{-\infty}^{\infty} d\tau_2\psi(\tau_1,\tau_2)\phi_j(\tau_2)=\lambda\phi_j(\tau_1) \tag{12}$$

may be solved, then e_7 is

$$e_7=\sum_{ij}\mu_i\mu_j\int_{-\infty}^{\infty} d\tau_1\phi_i(\tau_1)\lambda\phi_j(\tau_1), \tag{13}$$

but this integral will be zero unless $i=j$ because of the orthonormal properties of $\phi(\tau)$. Therefore

$$e_7=\sum_i \mu_i^2\lambda_i, \tag{14}$$

where the λ_i; are the eigenvalues of Eq. (12). Thus, e_7 has now been expressed as the sum of a number of independent random variables whose individual statistical properties are known, and therefore the method of characteristic functions may be applied.

2.31 Characteristic Function Method

It is already known from the discussion of Sec. 2.11 that the μ_n coefficients are normally distributed with zero mean and unity standard deviation. Therefore, the density of the random variable μ_n^2 is the well-known χ^2 distribution.[6] Using this density, the characteristic function or Fourier transform of the probability density for e_7 is

$$\Phi(\xi)=\int_0^{\infty} d(\lambda_1\mu_1)\exp(j\xi\lambda_1\mu_1^2)P(\lambda_1\mu_1^2) \cdots \int_0^{\infty} d(\lambda_n\mu_n^2)\exp(j\xi\lambda_n\mu_n^2)P(\lambda_n\mu_n^2). \tag{15}$$

Using the relation $P(\lambda_n\mu_n^2)d(\lambda_n\mu_n^2)=P(\mu_n^2)d(\mu_n^2)$ and the fact that all the integrations are independent, the

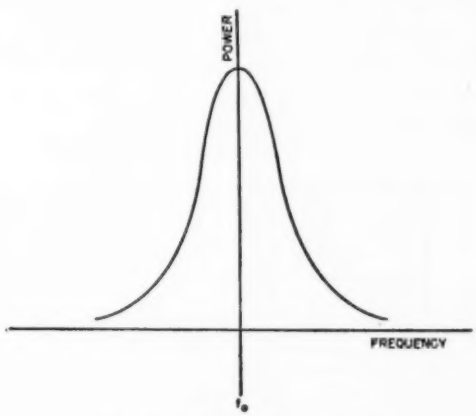


FIG. 2a. Input power spectrum.

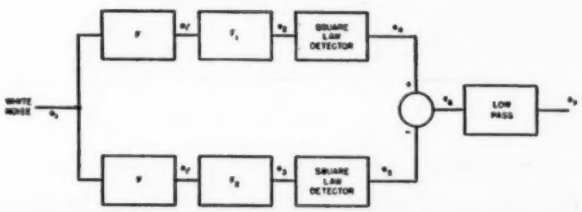


FIG. 2b. Block diagram of discriminator.

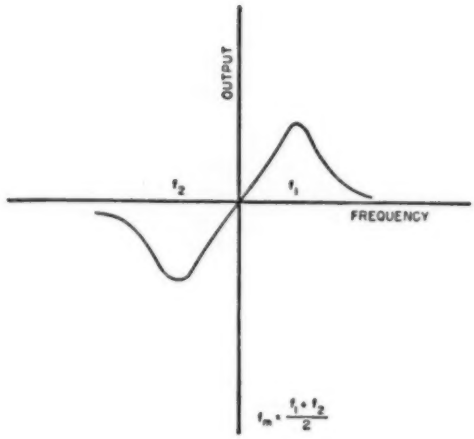


FIG. 3a. Discriminator characteristic.

characteristic function for e_7 becomes[8]

$$\Phi(\xi)=\frac{1}{\prod_{k=1}^{N}(1-2j\xi\lambda_K)^{\frac{1}{2}}}. \tag{16}$$

Thus, the probability density for e_7 is

$$P(e_7)de_7=\frac{de_7}{2\pi}\int_{-\infty}^{\infty}d\xi\exp(-j\xi e_7)\frac{1}{\prod_{K=1}^{N}(1-2j\xi\lambda_K)^{\frac{1}{2}}}. \tag{17}$$

Although it is probably possible to obtain a solution for $P(e_7)$, in general, it appears from the nature of the integral that it will be necessarily complicated. Thus, it will be more convenient to consider Eq. (17) as giving the general result for $P(e_7)$ and evaluating particular cases by substituting in the number of eigenvalues for that case and then integrating.

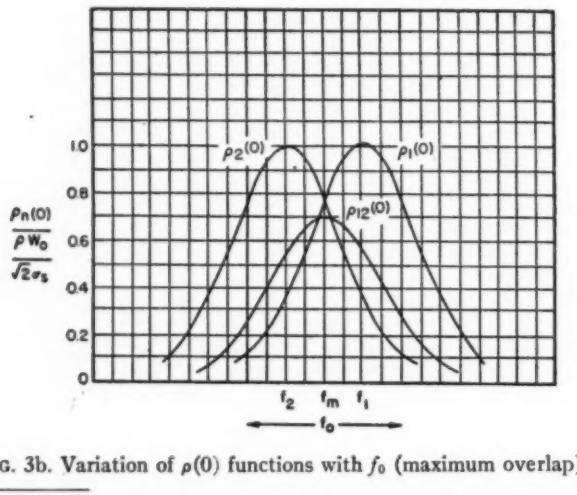


FIG. 3b. Variation of $\rho(0)$ functions with f_0 (maximum overlap).

[8] D. Bierens de Haan, *Novelles Tables D'Intégrales Definies* (G. E. Stechert and Company, New York, 1939).

3. FREQUENCY DISCRIMINATOR WITHOUT LOW PASS FILTER

3.1 Solution of Integral Equation

The expression for e_7 may be solved for no low pass filter by considering the following relation to be true.

$$\int_{-\infty}^{\infty}d\theta K(\theta)\gamma(\theta)=\gamma(0) \tag{18}$$

for all functions γ. This means that the weighting function, $K(\theta)$, is a delta function corresponding to infinite

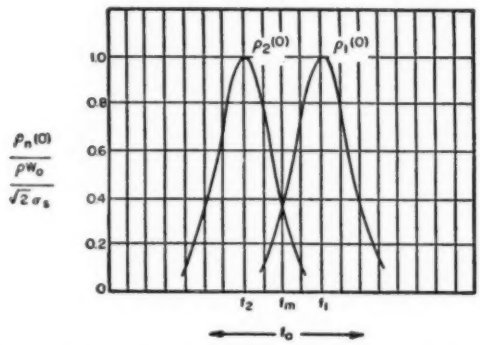


FIG. 3c. Variation of $\rho(0)$ functions with f_0 (negligible overlap).

low pass filter band width or, more simply, no filter at all. If Eq. (18) is true, the kernel of Eq. (12) become

$$\psi(\tau_1,\tau_2)=Q_1(\tau_1)Q_1(\tau_2)-Q_2(\tau_1)Q_2(\tau_2). \tag{19}$$

The eigenvalues for this kernel are two in number. The solution to Eq. (12) for the aforementioned kernel may

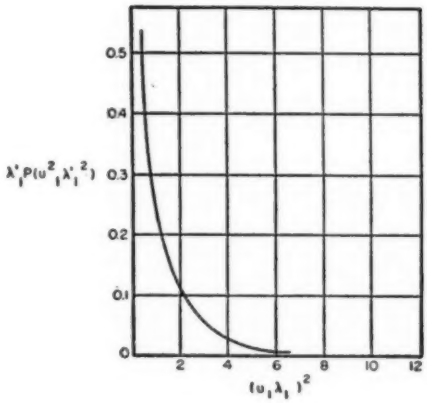


FIG. 3d. Probability density for $(\mu_1\lambda_1')^2$.

be found by standard methods for the solution of integral equations.[9,10] The values of the two eigenvalues are

[9] H. Margenau and G. M. Murphy, *The Mathematics of Physics and Chemistry* (D. Van Nostrand Company, Inc., New York, 1943).

[10] E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge University Press, Cambridge, 1950).

given below as

$$\left.\begin{matrix}\lambda_1\\ \lambda_2\end{matrix}\right\}=\frac{\rho_1-\rho_2}{2}\pm\tfrac{1}{2}[(\rho_1+\rho_2)^2-4\rho_{12}^2]^{\frac{1}{2}}, \qquad (20)\ddagger$$

where ρ_1 is the autocorrelation of e_2, ρ_2 the auto correlation of e_3, and ρ_{12} the cross correlation of e_2 and e_3, all with $\tau=0$. Observe that here the eigenvalue λ_1 will always be positive while the eigenvalue λ_2 will always be negative. Thus, the problem in this particular case is to find the probability density of the difference of the squares of two independent and normal random variables.

3.2 Filter Characteristics

It will be convenient at this point to introduce the nature of the filter characteristics so that when the probability density for no low pass filter is determined, the effects of filter variation and variation of input center frequency on that density may be observed.

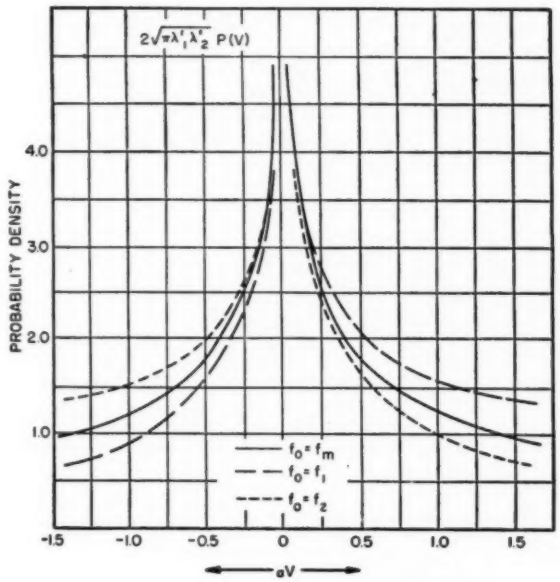


FIG. 3e. Output probability density for three values of f_0.

Now, in most physical systems the band pass filters consist of tuned RLC circuits. However, if the transfer of such a circuit be used in evaluating the correlation functions an extremely cumbersome integral results. Therefore, it will be most convenient to consider Gaussian filter characteristics since such characteristics may be approximated by cascaded filters, and also previous work in this field has shown statistical properties depend mostly on band width rather than actual filter shape. There is also the problem of a reasonable phase function. For the Gaussian amplitude function Bode's integral yields a divergent result. This is not as serious as one might expect since the phase factor appears only in the cross-correlation term. This cross-correlation due to filter overlap is, in general, sufficiently small to justify the assumption of linear phase shift with frequency. The filters F_1 and F_2 are taken with transfers of the form

$$F_n(f)=\exp\left[-\frac{(f-f_n)^2}{2\sigma^2}\right]\exp[-ja(f-f_n)]$$
$$+\exp\left[-\frac{(f+f_n)^2}{2\sigma^2}\right]\exp[-ja(f+f_n)], \qquad (21)$$

where a=slope of phase characteristic, f_n=center frequency and is f_1 for F_1 and f_2 for F_2, and σ=band width. This representation in Fig. 1b yields the discriminator characteristic of Fig. 3a.

The filter F is defined in such a fashion as to yield the desired Gaussian power spectrum. Therefore, one has

$$F(f)=\frac{1}{[2\sigma_s(2\pi)]^{\frac{1}{2}}}\left\{\exp\left[-\frac{(f-f_0)}{4\sigma_s^2}\right]^2\right.$$
$$\left.+\exp\left[-\frac{(f+f_0)^2}{4\sigma_s^2}\right]\right\}. \qquad (22)$$

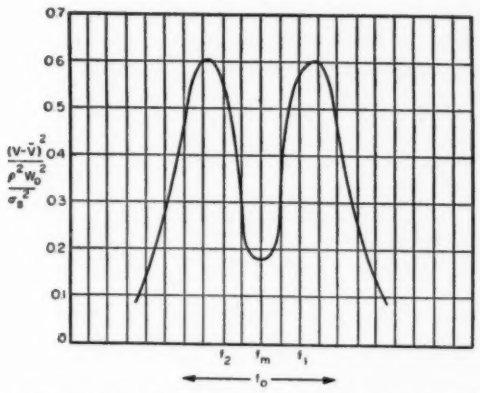


FIG. 3f. Variation of variance of V with no low pass filter (maximum overlap).

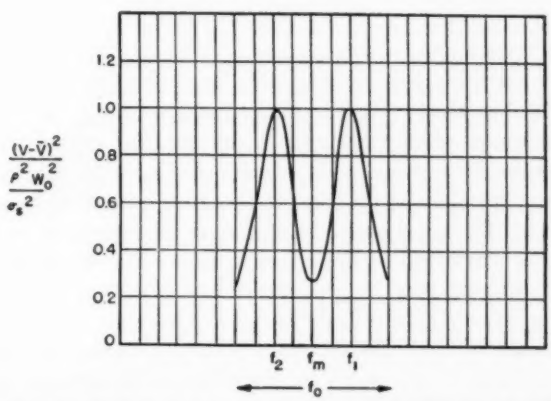


FIG. 3g. Variation of variance of V with no low pass filter (negligible overlap).

‡ Cross correlation due to filter overlap will always be such as to yield a real square root.

The power spectrum of the output of this filter is $W_0 \times |F(f)|^2$. Neglecting the cross-product terms, which are very small for the narrow band assumed, one has

$$W(f)=\frac{W_0}{(2\pi)^{\frac{1}{2}}2\sigma_s}\times\left\{\exp\left[-\frac{(f-f_0)^2}{2\sigma_s^2}\right]+\exp\left[-\frac{(f+f_0)^2}{2\sigma_s^2}\right]\right\} \quad (23)$$

as the power spectrum applied to the discriminator. The mean square value of this input is

$$\langle e_1'(t)^2\rangle_{Av}=\int_{-\infty}^{\infty} df W(f)=W_0. \quad (24)$$

Equation (24) implies the input signal power is constant regardless of its band width as long as that band width is narrow.

Now that F, F_1, and F_2 have been specified, the equivalent filter transfers FF_1 and FF_2 may be determined. This is done in the Appendix.

3.2.1 Variation of Functions with f_0

It is now convenient to examine the variation of the functions $\rho_1(\tau)$, $\rho_2(\tau)$, and $\rho_{12}(\tau)$ when τ equals zero as f_0 the center frequency of the input spectrum is varied. As will be seen shortly, it is these functions which determine the shape of the probability density and its moments when no low pass filter is present. The expressions for $\rho_1(\tau)$ and $\rho_2(\tau)$ may be evaluated using Eq. (6) and the Appendix results. The values for ρ_1 and ρ_2 at $\tau=0$ are

$$\rho_1(0)=\frac{\rho W_0}{\sqrt{2}\sigma_s}\exp\left[-\frac{(f_0-f_1)^2}{\sigma^2+2\sigma_s^2}\right],$$

$$\rho_{12}(0)=\frac{\rho W_0}{\sqrt{2}\sigma_s}\exp\left[-\frac{(f_1-f_2)^2}{4\sigma^2}\right]\exp\left[-\frac{(f_0-f_m)^2}{\sigma^2+2\sigma_s^2}\right], \quad (25)$$

$$\rho_2(0)=\frac{\rho W_0}{\sqrt{2}\sigma_s}\exp\left[-\frac{(f_0-f_2)^2}{\sigma^2+2\sigma_s^2}\right], \quad \rho^2=\frac{2\sigma^2\sigma_s^2}{\sigma^2+2\sigma_s^2}.$$

For particular applications, the value of the term $\exp[-(f_1-f_2)^2/4\sigma^2]$ may vary from a minimum value of 10^{-4} to about 0.7. The latter case, of course, will give maximum correlation and in general is due to filter overlap in the adjacent channels. Usually, the filter center frequencies are kept constant and their width σ is varied. Thus, the above variation may be due to varying σ. The variation of the ρ functions with f_0 is shown in Fig. 3b with maximum correlation. For this case, the ratio $(f_1-f_2)/(\sigma^2+2\sigma_s^2)^{\frac{1}{2}}$ is approximately unity. For the case of negligible overlap, this ratio becomes about two to one. The $\rho_1(0)$ and $\rho_2(0)$ are plotted for this condition in Fig. 3c. ρ_{12} is not plotted since it is negligibly small.

3.3 Probability Density for No Low-Pass Filter

From Eq. (20) it is seen that when no low pass filter is present the integral equation yields two eigenvalues, one positive and one negative. The probability density of the discriminator output V is from Eq. (17),

$$P(V)=\frac{1}{2\pi}\int_{-\infty}^{\infty} d\xi\epsilon^{-j\xi V}\frac{1}{[(1-2j\xi\lambda_1)(1-2j\xi\lambda_2)]^{\frac{1}{2}}}. \quad (26)$$

This integral may be evaluated by contour integration. However, since there are only two eigenvalues, the use of a real convolution integral is more convenient. By Eq. (14) the output V is

$$V=(\mu_1\lambda_1')^2-(\mu_2\lambda_2')^2, \quad (27)$$

where

$$(\lambda_1')^2=\lambda_1, \quad (\lambda_2')^2=-\lambda_2, \quad \lambda_1', \lambda_2'>0. \quad (28)$$

The problem now is to find the probability density of the difference of the squares of two Gaussian random variables $(\mu_1\lambda_1')$ and $(\mu_2\lambda_2')$ each with zero mean and standard deviation λ_1' and λ_2', respectively. The densities of $(\mu_1\lambda_1')^2$ and $-(\mu_2\lambda_2')^2$ are known. Thus the density of their sum is

$$P(V)=\frac{1}{2\pi(\lambda_1'\lambda_2')^{\frac{1}{2}}}\times\int dx\frac{\exp(-x/2\lambda_1')\exp(-V-x/2\lambda_2')}{\sqrt{x(x-V)^{\frac{1}{2}}}}, \quad (29)$$

where for $V>0$ the limits of integration are V to ∞, while for $V<0$ the limits are 0 to ∞. Consider first $V<0$. This gives

$$P(V)=\frac{\exp(V/2\lambda_2')}{2\pi(\lambda_1'\lambda_2')^{\frac{1}{2}}}\times\int_0^{\infty} dx\frac{\exp[-x(1/2\lambda_1'+1/2\lambda_2')]}{\sqrt{x(x-V)^{\frac{1}{2}}}}. \quad (29a)$$

This integral may be evaluated for V near zero by the following method. Whittaker[10] has considered a generalized hypergeometric function as defined by the following integral.

$$W_{k,m}(z)=\frac{z^k\epsilon^{-z/2}}{\Gamma(\frac{1}{2}-k+m)}\times\int_0^{\infty} dt t^{-k+\frac{1}{2}+m}\left(1+\frac{t}{z}\right)^{k-\frac{1}{2}+m}\epsilon^{-t}, \quad (30)$$

where $W_{k,m}(z)$ is the solution of the differential equation,

$$d^2W/dz^2+\left[-\tfrac{1}{4}+\frac{k}{z}+\frac{(\frac{1}{4}-m^2)}{z^2}\right]W=0. \quad (31)$$

This equation has a regular singularity at zero and an irregular one at infinity. If now $k=m=0$, Eq. (30) becomes

$$W_{0,0}(z)=\frac{\epsilon^{-z/2}}{\Gamma(\frac{1}{2})}\int_0^\infty dt\frac{\epsilon^{-t}}{t^{\frac{1}{2}}(1+t/z)^{\frac{1}{2}}}. \tag{32}$$

This may be used to find an expression for $P(V)$ with V near zero which is the region of interest, by solving Eq. (31) for a value in the vicinity of its regular singularity. For large values of V an asymptotic expression for W is most reasonable. If (31) is solved with $k=m=0$, for z near zero, and this result and Eqs. (29a) and (32) are used, one has for $V<0$

$$P(V)=\frac{\exp(v/2\lambda_2')\exp(-\alpha v/2)}{2(\pi\lambda_1'\lambda_2')^{\frac{1}{2}}}$$
$$\times\left\{\left[1+\frac{(\alpha v)^2}{16}+\frac{(\alpha v)^4}{1024}+\cdots\right][R_1+R_2\log|\alpha v|]\right.$$
$$\left.-R_2\left[\frac{(\alpha v)^2}{16}+\frac{3(\alpha v)^4}{2048}+\cdots\right]\right\},$$

where

$$\alpha=\frac{\lambda_1'+\lambda_2'}{2\lambda_1'\lambda_2'}. \tag{33}$$

Since $P(V)$ for $V=0$ tends to infinity in logarithmic fashion, R_2 in (33) is chosen as -1. R_1 may be chosen so $P(V)>0$ over the range of interest. In other words, it exhibits the properties of a probability density. In a similar fashion for $V>0$ the density is

$$P(V)=\frac{\exp(-v/2\lambda_1')\exp(\alpha v/2)}{2(\pi\lambda_1'\lambda_2')^{\frac{1}{2}}}\left\{\left[1+\frac{(\alpha v)^2}{16}+\cdots\right]\right.$$
$$\left.\times\left[A+\log\frac{1}{\alpha v}\right]+\frac{(\alpha v)^2}{16}+\frac{3(\alpha v)^4}{2048}+\cdots\right\}. \tag{34}$$

In Fig. 3d, the probability density for $(\mu_1\lambda_1')^2$ is shown. That for $-(\mu_2\lambda_2')^2$ has the same form. Figure 3e shows the probability density of V for various values of f_0. Observe that for $f_0=f_m$ the density is a symmetrical one with zero mean, which is entirely reasonable in light of the physical aspects of the problem. As f_0 goes to either side of f_m, the output of one channel is large and the other small. This causes the density to be assymetrical. The plot of Fig. 3e is for maximum overlap and gives two extreme cases $f_0=f_1$ and $f_0=f_2$. From Fig. 3b this gives a ratio of λ_1'/λ_2' of 2.72 and 1/2.72, respectively. For f_0 at other points between f_1 and f_2 the density curves will fall between these extremes. In the case of negligible overlap with ρ_{12} approximately zero, these same extreme values of f_0 give essentially no output in one channel and thus the value of one is approximately zero and the other is $\sqrt{\rho}$ for the channel with output. Here no convolution integral is needed and $P(V)$ becomes that of Fig. 3d for $f_0=f_1$ with negligible overlap. For f_0 a great distance from f_m, the problem reduces itself to a single channel and again the probability density is also of the form of Fig. 3d.

3.31 Moments of the Density Function

The mean value of V is the average of Eq. (28) and since $\langle\mu_1^2\rangle=\langle\mu_2^2\rangle_{Av}=1$, this becomes

$$\bar{V}=\lambda_1'^2-\lambda_2'^2$$
$$=\rho_1(0)-\rho_2(0). \tag{35}$$

In terms of the filter transfers, this is

$$\bar{V}=\frac{\rho W_0}{\sqrt{2}\sigma_s}\left\{\exp\left[-\frac{(f_0-f_1)^2}{\sigma^2+2\sigma_s^2}\right]-\exp\left[-\frac{(f_0-f_2)^2}{\sigma^2+2\sigma_s^2}\right]\right\}. \tag{36}$$

Equation (36) is essentially the frequency discriminator characteristic drawn in Fig. 3a which give a measure of dc output as f_0 varies. The variance of the output is

$$\text{Variance of } V=\langle V^2\rangle_{Av}-(\bar{V})^2$$
$$=\langle[(\mu_1\lambda_1')^2-(\mu_2\lambda_2')^2]^2\rangle_{Av}-\bar{V}^2. \tag{37}$$

The variance in terms of the correlation functions is

$$\langle(V-\bar{V})^2\rangle_{Av}=\frac{\rho^2W_0^2}{\sigma_s^2}\left\{\exp\left[-\frac{2(f_0-f_1)^2}{\sigma^2+2\sigma_s^2}\right]\right.$$
$$+\exp\left[-\frac{2(f_0-f_2)^2}{\sigma^2+2\sigma_s^2}\right]-2\exp\left[-\frac{(f_1-f_2)^2}{2\sigma^2}\right]$$
$$\left.\times\exp\left[-\frac{2(f_0-f_m)^2}{\sigma^2+2\sigma_s^2}\right]\right\}. \tag{38}$$

The variance as a function of f_0 is plotted in Fig. 3f. From this it can be seen that the spread of the density is a minimum when f_0 is centered between the two filters, which is as expected. As the frequency departs radically to the outside of f_1 or f_2 the variance approaches either $2\rho_1^2(0)$ or $2\rho_2^2(0)$ which, of course, eventually go to zero when f_0 falls a considerable distance outside the band, since then neither channel has an output. Observe also that the variance rises quite sharply as f_0 departs from f_m. If there is no overlap or negligible overlap, the variance reduces to the first two terms of Eq. (38). This variation is plotted in Fig. 3g. It is considerably steeper in its variation than the case for overlap.

3.4 Conclusions for the Case of No Filter

In summing up this portion of the probability problem, it may be said that the density of the difference of the squares of two independent Gaussian variables exhibits similar properties to its component densities. In this

physical problem, these independent components do not correspond to the voltage in each individual channel but are actually combinations of those voltages divided in such a fashion as to give two independent random variables. The moments of the density may be found by averaging, using the statistical properties of the μ_n quantities. However, it will be shown in the following sections that whenever the voltage is expressed as in Eq. (14), the moments may be found from

$$n=\sum_1^N \lambda_j{}^n$$

where n is the order of the moment, provided the kernel of the integral equation is symmetric. This will be especially useful in evaluating the properties of the probability density after the low pass filter.

4. FREQUENCY DISCRIMINATOR WITH LOW PASS FILTER

4.1 Method of Moments

As was shown in the preceding sections the integral equation when no low pass filter is present is rather simple to solve. However, when $K(\theta)$ is not a delta function, the integral equation becomes quite cumbersome and another approach is more fruitful. If some method may be used by which the significant moments at the output of the low pass filter might be found, some useful information concerning the output probability density is available. For example, its approach to a Gaussian density as a function of band width ratios becomes very clear.

From Eq. (16) the characteristic function of the low pass filter output was found to be

$$\Phi(\xi)=\frac{1}{\prod_{K=1}^N (1-2j\xi\lambda_K)^{\frac{1}{2}}}. \tag{16}$$

If one takes the logarithm of this quantity and expands it in a series of terms of the form $(j\xi)^n$ one has

$$\log\Phi(\xi)=\tfrac{1}{2}\left[\sum_1^N 2\lambda_K(j\xi)+\tfrac{1}{2}\sum_1^N 4\lambda_K{}^2(j\xi)^2 + \cdots \frac{1}{n}\sum_1^N 2^n\lambda_K{}^n(j\xi)^n+\cdots\right]. \tag{39}$$

This expansion for $\log\Phi(\xi)$ is used frequently in probability and statistics. The coefficients in this expansion are given the names semi-invariants of the probability density corresponding to $\Phi(\xi)$. These semi-invariants bear definite relations to the moments of the density.[11] If these coefficients are denoted by χ_n, one has

$$\log\Phi(\xi)=\sum_1^N \frac{\chi_K}{K!}(j\xi)^K. \tag{40}$$

When these coefficients are compared to those in the expansion of Eq. (39) one has

$$\begin{aligned} &\sum_1^N \lambda_K=\chi_1=m, \quad 8\sum_1^N \lambda_K{}^3=\chi_3=\mu_3, \\ &2\sum_1^N \lambda_K{}^2=\chi_2=\mu_2, \quad 48\sum_1^N \lambda_K{}^4=\chi_4=\mu_4-3\mu_2{}^2, \end{aligned} \tag{41}$$

and so on, where μ_n is the nth central moment about the mean of the density, with that mean value denoted by m. The problem, therefore, is to find some way of evaluating the eigenvalue sums in terms of the weighting and correlation functions of the system in question. This has been done previously.[11,12] Therefore, the successive sums may be evaluated by using the n times iterated kernel of Eq. (12). For higher moments this evaluation by integration becomes considerably involved, but actually it is not unreasonable to obtain the first three semi-invariants and thus the first three moments using this method and also an indication of the nature of the fourth semi-invariant. This will be accomplished in the following sections.

4.2 Eigenvalue Sums in Terms of System Transfers

Consider now the kernel of the integral equation arising in the case of the frequency discriminator with low pass filter. Using the iterated kernel with n equal to unity,

$$\int_{-\infty}^{\infty} dx\psi(x,x)=\int_{-\infty}^{\infty} dx\int_{-\infty}^{\infty} d\theta K(\theta) \times[Q_1{}^2(x-\theta)-Q_2{}^2(x-\theta)]=\sum\lambda_i. \tag{42}$$

In terms of the correlation functions, this becomes

$$\sum\lambda_i=m=\int_{-\infty}^{\infty} d\theta K(\theta)[\rho_1(0)-\rho_2(0)]. \tag{43}$$

In a similar fashion:

$$\sum\lambda_i{}^2=\int_{-\infty}^{\infty} d\theta K(\theta)\int_{-\infty}^{\infty} d\beta K(\beta)[\rho_1{}^2(\beta-\theta) +\rho_2{}^2(\beta-\theta)-\rho_{12}{}^2(\beta-\theta)-\rho_{21}{}^2(\beta-\theta)] \tag{44}$$

[11] It has come to the author's attention that this approach had been developed for AM signals at about the same time this paper was originally written. See R. C. Emerson, J. Appl. Phys. **24**, 1168 (1953).

[12] G. R. Arthur, J. Appl. Phys. **23**, 1143 (1952).

and

$$\sum \lambda_i^3 = \int_{\infty}^{\infty} d\theta K(\theta) \int_{-\infty}^{\infty} d\beta K(\beta) \times \int_{-\infty}^{\infty} d\gamma K(\gamma) [\rho_1(\gamma-\theta)\rho_1(\beta-\theta)\rho_1(\gamma-\beta) - \rho_2(\gamma-\theta)\rho_2(\beta-\theta)\rho_2(\gamma-\beta) - \rho_{12}(\gamma-\theta)\rho_1(\beta-\theta)\rho_{12}(\gamma-\beta) + \rho_{21}(\gamma-\theta)\rho_2(\beta-\theta)\rho_{21}(\gamma-\beta) - \rho_1(\gamma-\theta)\rho_{12}(\beta-\theta)\rho_{21}(\gamma-\beta) + \rho_{12}(\gamma-\theta)\rho_{12}(\beta-\theta)\rho_2(\gamma-\beta) - \rho_{21}(\gamma-\theta)\rho_{21}(\beta-\theta)\rho_1(\gamma-\beta) + \rho_2(\gamma-\theta)\rho_{21}(\beta-\theta)\rho_{12}(\gamma-\beta)]. \quad (45)$$

During the integration, it is recognized that some of the terms in Eq. (45) yield negligible results so the equation is not as formidable as one might expect.

4.3 Evaluation of Moments

Before evaluating the semi-invariants, it is necessary to specify the low pass filter weighting function. Equation (43) shows that the mean of the output density is that of the input modified by the factor

$$\int_{-\infty}^{\infty} d\theta K(\theta).$$

Therefore, it will be most convenient to make a normalization such that

$$\int_{-\infty}^{\infty} d\theta K(\theta) = 1 \quad (46)$$

and therefore the mean values on both sides of the low pass filter are the same. Thus the low pass device will be characterized by a weighting function

$$K(t) = (2\pi)^{\frac{1}{2}}\sigma_l \exp(-2\pi^2\sigma_l^2 t^2) = K \exp(-2\pi^2\sigma_l^2 t^2). \quad (47)$$

This means the low pass filter is Gaussian with standard deviation σ_l. Again, an RC filter is more reasonable physically, while the Gaussian is convenient mathematically. Since the statistics depend mainly on band width, the assumption of a Gaussian filter is not restrictive (see Sec. 3.2).

Therefore, the first semi-invariant, and by Eq. (41) the mean of the probability density after the low pass filter is

$$\langle e_7 \rangle_{Av} = \frac{\rho W_0}{\sqrt{2}\sigma_s}\left\{\exp\left[-\frac{(f_0-f_1)^2}{\sigma^2+2\sigma_s^2}\right] - \exp\left[-\frac{(f_0-f_2)^2}{\sigma^2+2\sigma_s^2}\right]\right\}, \quad (48)$$

which is identical with the input mean due to the normalization. The second semi-invariant may be evaluated from Eq. (44). If this is done, the integration of the first term yields

$$\sum \lambda_i^2 = \tfrac{1}{2}\frac{\sigma_l^2}{(\sigma_l^4+2\sigma_l^2\rho^2)^{\frac{1}{2}}}(A_1^2+B_1^2-2C_1^2) + \text{integrals in } \cos\beta, \quad (49)$$

where

$$A_1 = \frac{\rho W_0}{\sqrt{2}\sigma_s}\exp\left[-\frac{(f_0-f_1)^2}{\sigma^2+2\sigma_s^2}\right],$$

$$B_1 = \frac{\rho W_0}{\sqrt{2}\sigma_s}\exp\left[-\frac{(f_0-f_2)^2}{\sigma^2+2\sigma_s^2}\right],$$

$$C_1 = \frac{\rho W_0}{\sqrt{2}\sigma_s}\exp\left[-\frac{(f_0-f_m)^2}{\sigma^2+2\sigma_s^2}\right]\exp\left[-\frac{(f_1-f_2)^2}{4\sigma^2}\right].$$

If the cosine terms are integrated one obtains a multiplying factor of the form

$$\exp\left(-\frac{2f_1'^2}{\rho^2+\sigma_l^2}\right).$$

Now in this investigation, only narrow band systems are being discussed and in a particular application, the value of the exponent is of the order of 100 or more. This means each term is multiplied by a factor ϵ^{-100} which is indeed negligible. The remaining integrations in β yield the same type of term. If the integration is carried through with respect to θ one still obtains a factor which for large values of low pass band width does not exceed $\epsilon^{-20} = 25\times10^{-6}$ which is a negligible factor. Therefore, the variance of the low pass filter output becomes

$$2\sum \lambda_i^2 = \mu_2 = \frac{\sigma_l^2}{(\sigma_l^4+2\sigma_l^2\rho^2)^{\frac{1}{2}}}[A_1^2+B_1^2-2C_1^2]. \quad (50)$$

Substituting for A_1, B_1, and C_1 with $\rho/\sigma_l > 5$,

$$\mu_2 = \frac{\sigma_l \rho W_0^2}{2\sqrt{2}\sigma_s^2}\left\{\exp\left[-\frac{2(f_0-f_1)^2}{\sigma^2+2\sigma_s^2}\right]+\exp\left[-\frac{2(f_0-f_2)^2}{\sigma^2+2\sigma_s^2}\right] - 2\exp\left[-\frac{2(f_0-f_m)^2}{\sigma^2+2\sigma_s^2}\right]\exp\left[-\frac{(f_1-f_2)^2}{2\sigma^2}\right]\right\}. \quad (51)$$

The variation of variance from input to output with respect to f_0 is unchanged as far as shape is concerned. The ratio of variance at the output to that at the input is

$$\frac{\text{Output variance}}{\text{Input variance}} = \frac{\sigma_l}{2\sqrt{2}\rho}. \quad (52)$$

This, in general, will be a very small quantity.

One may now proceed to the evaluation of the third semi-invariant or third central moment and observe its relation to μ_2. This will give an indication of the manner in which the output probability distribution approaches a Gaussian one when the band of the low pass device becomes narrow. Using Eq. (45) for the $\sum \lambda_i^3$, it can be shown that only two of the terms yield a non-negligible

result. These are the first two terms of Eq. (45) which contain the ρ_1 and ρ_2 functions alone.

An extensive investigation of all of Eq. (45) has indicated that the cross-correlation has little effect on the third moment. This is reasonable when one remembers that the cross-correlation arises from common components due to filter overlap and when the outputs of each channel are subtracted, these tend to cancel. The odd-order moments retain some of the cancellation characteristic, and therefore cross-correlation has a small effect on μ_3.

The result, therefore is that for an effective "signal" band width five or more times the low pass band width one has

$$\mu_3 \approx 8 \sum \lambda_i^3 = \frac{2\rho W_0^3 \sigma_i^2}{3\sqrt{2}\sigma_s^3} \times \left\{ \exp\left[-\frac{3(f_0-f_1)^2}{\sigma^2+2\sigma_s^2} \right] - \exp\left[-\frac{3(f_0-f_2)^2}{\sigma^2+2\sigma_s^2} \right] \right\}. \quad (53)$$

This gives a variation of μ_3 with f_0 similar to Fig. 3a, but with a narrower and steeper characteristic. This is shown in Fig. 4a where normalized values of μ_1 and μ_3 are plotted for comparison as a function of f_0. Normalization is such that only the bracketed term is plotted. Also, for $f_0 = f_m$ both μ_1 and μ_3 are negative, but in 4a only magnitude is plotted.

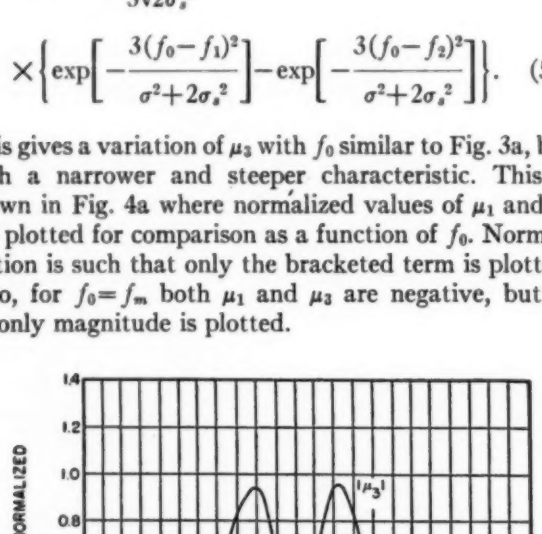


FIG. 4a. Variation of $|\mu_1|$ and $|\mu_3|$ with f_0.

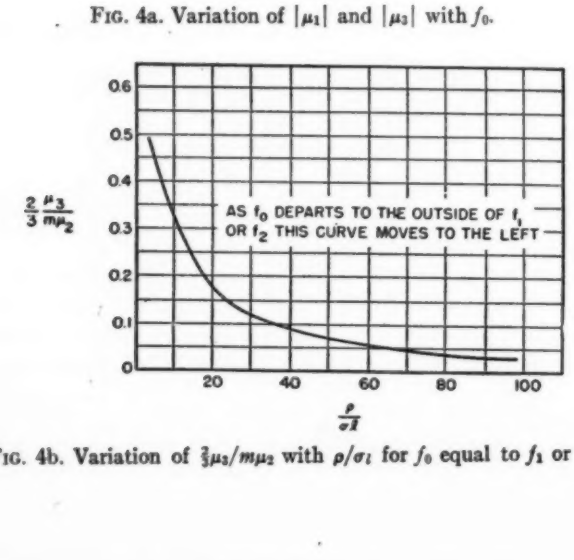


FIG. 4b. Variation of $\frac{2}{3}\mu_3/m\mu_2$ with ρ/σ_I for f_0 equal to f_1 or f_2.

In evaluating the fourth semi-invariant, the expression for $\sum \lambda_i^4$ appears quite cumbersome. However, some inferences concerning the nature of this coefficient may be drawn by observing that all the semi-invariants so far evaluated received a dominant contribution from the terms involving $\rho_1(\tau)$ alone and $\rho_2(\tau)$ alone. This is apparent also from physical considerations of the problem. Thus, one could draw some conclusions by evaluating the first term in the integration in the equation for $\sum \lambda_i^4$. If this is done, it is found that the fourth semi-invariant is near zero. Therefore, one can examine the approach of the output probability density to a Gaussian density by observing the ratio of the third and second semi-invariants.

4.4 Approach of Density to Gaussian[12]

A Gaussian probability density is given by

$$P(x) = \frac{1}{(2\pi)^{\frac{1}{2}}\mu_2^{\frac{1}{2}}} \exp\left[-\frac{(x-m)^2}{2\mu_2} \right], \quad (54)$$

where m is the mean and μ_2 the variance. The characteristic function is

$$\Phi(u) = \exp(-\tfrac{1}{2}\mu_2 u^2) \exp(-jmu). \quad (55)$$

From this it may be seen that the only semi-invariants which exist are χ_1 and χ_2. Consequently, a density in which these predominate is close to a Gaussian since the characteristic function uniquely determines a probability density. If the effect of the third semi-invariant, which is also the third central moment, is to be considered, the characteristic function is

$$\Phi(u) = \exp(-\tfrac{1}{2}\mu_2 u^2) \exp\left[-j\left(mu + \frac{\mu_3}{6}u^3 \right) \right]. \quad (56)$$

The magnitude term here is Gaussian and if the term denoting the argument remains almost linear over the range of u where

$$|\Phi(u)| \geq \epsilon^{-2},$$

the Fourier transform of $\Phi(u)$, which is the probability density, will be nearly Gaussian. Thus, the u value of interest is

$$u^2 = 4/\mu_2.$$

For this the argument becomes

$$\operatorname{Arg}\Phi(u) = \exp\left(-j\frac{2m}{\mu_2^{\frac{1}{2}}} \right) \left\{ 1 + \tfrac{2}{3}\frac{\mu_3}{m\mu_2} \right\}. \quad (57)$$

Consider the ratio $\frac{2}{3}\mu_3/m\mu_2$. This is

$$\frac{2}{3}\frac{\mu_3}{m\mu_2} = \frac{4\sqrt{2}\sigma_I}{3\rho} \times \left\{ \frac{A_2^2+B_2^2+A_2B_2}{A_2^2+B_2^2-2(A_2B_2)^{\frac{1}{2}} \exp\left[-\frac{\sigma_s^2}{\sigma^2}\frac{(f_1-f_2)^2}{\sigma^2+2\sigma_s^2} \right]} \right\}, \quad (58)$$

where

$$A_2=\exp\left[-\frac{(f_0-f_1)^2}{\sigma^2+2\sigma_s^2}\right], \quad B_2=\exp\left[-\frac{(f_0-f_2)^2}{\sigma^2+2\sigma_s^2}\right].$$

The ratio of the bracketed terms is essentially the ratio of

$$\mu_3'/m'\mu_2',$$

where the primes indicate the normalized values of these various means which have been plotted in Figs. 3f and 4a. Thus from these plots, it is possible to obtain values of their ratio. For wide low pass band widths, skewness or asymmetry is introduced in the density and only the narrowing of the low pass band will give a Gaussian density. For f_0 equal to f_2 or f_1, or outside these frequencies take the maximum of the variational component of Eq. (58). For this condition, the variation of $\frac{2}{3}\mu_3/m\mu_2$ is shown in Fig. 4b as a function of ρ/σ_l. For ρ/σ_l greater than 70 this factor is less than 0.05, and therefore, it seems reasonable that the output density will closely approximate a Gaussian one under these conditions. The equivalent filter band width is ρ and for most applications, the discriminator filters are broad enough so that from Eq. (25) one may conclude ρ is approximately the band width of the input spectrum. Actually the figure of 70 is for $f_0=f_1$ or f_2. As f_0 moves beyond these frequencies, lesser values of ρ/σ_l will give less than 5 percent contribution to the phase term and thus make the density near a Gaussian one as has been shown.

4.5 Conclusions

When the frequency discriminator is followed by a low pass filter, the predominant semi-invariants of the probability density are the first three.§ The fourth is near zero. Thus, the predominant moments of the density are the mean, the variance, the third moment about the mean. The fourth moment about the mean is approximately $3\mu_2^2$ found by assuming χ_4, equal to zero and any higher-order moments may be found by assuming all semi-invariants beyond the fourth are zero. The effect of the third moment on the density is to give it a skewness similar to that shown in Fig. 3e for the case of no filter. As the band width of the low pass filter is reduced, the effect of this third moment becomes progressively less and the density approaches the symmetrical Gaussian one. The degree to which the band of the output must be reduced depends on the location of f_0, the center frequency of the input spectrum relative to f_1 and f_2, the center frequencies of the two discriminator filters. This narrowness requirement becomes less severe as f_0 moves farther away from f_m and in the preceding section, the figure 1/70 has been given as the ratio of output band to signal band when f_0 is f_1 or f_2. For a particular application where the ratio of low pass band to "signal" band is 0.001 or less, it may be safely assumed that the output probability density is Gaussian regardless of the location of the input spectrum.

§ Low pass band width subject to restrictions of Eq. (51).

APPENDIX. PRODUCT OF TWO GAUSSIAN FACTORS

Consider the product of two Gaussian factors as in the following:

$$Z=\exp\left[-\frac{(x-f_0)^2}{2\gamma^2}\right]\exp\left[-\frac{(x-f_1)^2}{2\gamma_1^2}\right]. \quad \text{(A-1)}$$

This is

$$Z=\exp\left[-\frac{\gamma_1^2(x-f_0)^2+\gamma^2(x-f_1)^2}{2\gamma^2\gamma_1^2}\right]. \quad \text{(A-2)}$$

Let

$$\rho^2=\frac{\gamma^2\gamma_1^2}{\gamma^2+\gamma_1^2}.$$

Then (AB-2) is

$$Z=\exp\left[-\frac{1}{2\rho^2}\left\{\left(x-\frac{\gamma^2f_1+\gamma_1^2f_0}{\gamma^2+\gamma_1^2}\right)^2-\left(\frac{\gamma^2f_1+\gamma_1^2f_0}{\gamma^2+\gamma_1^2}\right)^2+\frac{f_0^2\gamma_1^2+\gamma^2f_1^2}{\gamma^2+\gamma_1^2}\right\}\right] \quad \text{(A-3)}$$

$$=\exp\left[-\frac{1}{2\rho^2}(x-f_0{}^*f_1)^2\right]\exp\left[-\tfrac{1}{2}\left\{\left(\frac{f_0}{\gamma}\right)^2+\left(\frac{f_1}{\gamma_1}\right)^2-\frac{1}{\gamma^2+\gamma_1^2}\left[\left(\frac{f_1\gamma}{\gamma_1}\right)^2+\left(\frac{f_0\gamma_1}{\gamma}\right)^2\right]\right\}\right]. \quad \text{(A-4)}$$

This becomes finally

$$Z=\exp\left[-\frac{1}{2\rho^2}(x-f_0{}^*f_1)^2\right]\exp\left[-\frac{(f_0-f_1)^2}{2(\gamma^2+\gamma_1^2)}\right], \quad \text{(A-5)}$$

where

$$f_0{}^*f_1=\frac{\gamma^2f_1+\gamma_1^2f_0}{\gamma^2+\gamma_1^2}.$$

This relation has been used throughout to simplify Gaussian products and yields for the general filter transfer function of the form

$$FF_n(f)=A\left[\exp\left[-\frac{(f-f_n')^2}{2\rho^2}\right]\exp[-ja(f-f_n)]+\exp\left[-\frac{(f+f_n')^2}{2\rho^2}\right]\exp[-ja(f+f_n)]\right], \quad \text{(A-6)}$$

where

$$A=\frac{1}{[2\sigma_s(2\pi)^{\frac{1}{2}}]^{\frac{1}{2}}}\exp\left[-\frac{(f_n-f_0)^2}{2(\sigma^2+2\sigma_s^2)}\right]$$

and

$$\rho^2=\frac{2\sigma^2\sigma_s^2}{\sigma^2+2\sigma_s^2}, \quad f_n'=\frac{\sigma^2f_0+2\sigma_s^2f_n}{\sigma^2+2\sigma_s^2}.$$

Amplification Factor and Perveance of an Elliptic Triode

S. DEB AND G. S. SANYAL
Institute of Radio Physics and Electronics, University of Calcutta, India
(Received December 4, 1953)

An attempt has been made to derive, in terms of valve geometry, theoretical expressions for the constants of an elliptic triode, the shape of the electrodes of which belongs to a family of confocal ellipses. The amplification factor (μ) is derived by following the method of conformal transformation. Result shows that in general it varies with the parametric angle θ of the ellipse—decreasing from the direction of minor to that of the major axis. The average value of μ increases with increasing value of the eccentricity of the grid and with decreasing value of the eccentricity of the anode. Curves for ready evaluation of μ for any value of θ are given. The problem of design of an elliptic triode having a constant μ for all values of θ is also considered and its method of solution indicated. An expression for the perveance (P) has also been obtained by adopting the procedure followed by O'Neill for determining the perveance of a plane diode having filamentary cathode. The expression shows that the value of perveance is determined mainly by the eccentricity of the grid and the focal distance.

To illustrate the practical applications of the various deductions made, a numerical example, involving the design of a 6C5-GT/G type triode, is given.

1. INTRODUCTION

A KNOWLEDGE, in terms of the electrode geometry, of the amplification factor, the interelectrode capacitances, and the perveance of an electronic valve is of great importance for the design and manufacture such valves. Many investigations have, therefore, been carried out to derive theoretical expressions for these parameters. The cases studied most are those of triodes and multigrid valves having planar and cylindrical geometry. Expressions are now available for calculating theoretically the various constants of these types of triodes with arbitrary electrode dimensions and spacings.[1–3] However, many commercial triodes have a geometry which is neither plane nor cylindrical. Of these, those possessing approximately an elliptic geometry are perhaps the most numerous. But it appears that no theoretical analysis has yet been attempted to obtain expressions for the constants of such triodes. In the existing design practice, therefore, such valves are treated more or less as of cylindrical or planar structure having electrode distances equal to the average distances of the valve electrodes, use being made of certain empirical relations for estimating the valve constants. Such a procedure is obviously unsatisfactory. In the present paper an attempt will be made to develop expressions for the amplification factor, interelectrode capacitances, and perveance of a triode of elliptic geometry. In order that the problem may be amenable to mathematical treatment, the case considered will be that for which the electrode surfaces constitute a system of confocal ellipses. Subject to certain reasonable restrictions, the analysis will be shown to yield expressions useful for practical valve design.

[1] K. R. Spangenberg, *Vacuum Tubes* (McGraw-Hill Book Company, Inc., New York, 1948).
[2] W. B. Walker, Proc. Inst. Elec. Engrs. (London), **98(III)**, 57 (1951).
[3] M. Wada, Scientific Rept. Research Insts. Tohoku Univ. **B-1, 2**, 399 (1950).

2. TRIODE OF ELLIPTIC GEOMETRY

As mentioned earlier, the special case considered will be that of a valve for which the cathode, the grid, and the anode surfaces are confocal elliptic cylinders. Let the eccentricities of these be e_c, e_g, and e_a, respectively. The grid is supposed to be made of straight wires, all of the same radius r_g, and disposed parallel to the common axis of the cylinders at equal intervals over the grid surface. A transverse section of the above structure consisting of the anode (A), the grid (G), and the cathode (C) is shown in Fig. 1. The directions of the axes and the center of the confocal elliptic system have been taken, respectively, as the directions and the origin of a rectangular coordinate system (x,y). The two foci of the ellipses are the points $(\pm K,0)$ on the (x,y) coordinate system. The centers of the grid wires divide the elliptic grid section into a number of equal arcs. The length D of each of these arcs is related to the total number of grid wires ($2N$) by

$$D=2KE/(Ne_g), \tag{1}$$

where $\delta_g=\sin^{-1}e_g$ and $E=E(\delta_g,\pi/2)$ represents the complete elliptic integral of the second kind.

3. CONFORMAL TRANSFORMATION OF THE TRIODE GEOMETRY

In order to investigate analytically the foregoing problem, it is necessary first to reduce the elliptic triode to a simpler geometry for which solutions are easily obtainable. For this purpose, we restrict our interest to the region of the triode lying above the x axis (Fig. 1) and apply the transformation given by

$$Z=K\sin(W/K), \tag{2}$$

where

$$\left.\begin{aligned} W&=\varphi+j\psi \\ Z&=x+jy \end{aligned}\right\}. \tag{3}$$

and

From Eqs. (2) and (3), we obtain

$$x=K\cosh(\psi/K)\sin(\varphi/K) \tag{4a}$$

$$y=K\sinh(\psi/K)\cos(\varphi/K). \tag{4b}$$

As a result of the aforementioned transformation, a system of confocal ellipses with their foci located at $(\pm K, 0)$ in the Z plane transform in the W plane into a series of lines parallel to the φ axis. It follows, therefore, that the elliptic arc EAF representing the upper half of the anode in Fig. 1 becomes in the W plane a straight line given by

$$\psi=K\cosh^{-1}(a_a/K)$$
$$=K\cosh^{-1}(1/e_a)=\psi_a, \text{ say} \tag{5}$$

where a_a is the semimajor axis of the elliptic anode A.

Putting $e_a=\sin\delta_a$ and using the identity $\cosh^{-1}u=\ln[u+(u^2-1)^{\frac{1}{2}}]$, Eq. (5) reduces to

$$\psi=\psi_a=K\ln\cot(\delta_a/2). \tag{6}$$

Similarly, the transformed grid and cathode arcs are the straight lines, given, respectively, by

$$\psi=K\ln\cot(\delta_g/2)=\psi_g, \text{ say} \tag{7}$$

and

$$\psi=K\ln\cot(\delta_c/2)=\psi_c, \text{ say} \tag{8}$$

where $\delta_c=\sin^{-1}e_c$.

It also follows from Eqs. (4a) and (4b) that the three transformed lines would extend from $+K\pi/2$ to $-K\pi/2$. The transformed electrode surfaces constitute a plane triode having the following electrode spacings:

Cathode-to-grid distance d_{cg}

$$=\psi_g-\psi_c=K\ln[\cot(\delta_g/2)/\cot(\delta_c/2)]. \tag{9}$$

Grid-to-anode distance d_{ga}

$$=\psi_a-\psi_g=K\ln[\cot(\delta_a/2)/\cot(\delta_g/2)]. \tag{10}$$

If, now, Eqs. (4a) and (4b) are compared with the parametric equation of an ellipse, *viz.* $x=a\sin\theta$, $y=b\cos\theta$, a and b being the semimajor and semiminor diameter, respectively, it is seen that

$$\theta=\varphi/K. \tag{11}$$

Equation (11) shows that points on any of the confocal ellipses which correspond to equal intervals of the angle θ in the Z plane are spaced at equal intervals of φ in the W plane. Now, equal arc lengths of an ellipse correspond to increasing intervals in θ as one goes away from the semiminor axis. It, therefore, follows from Eq. (11) that the spacings of the transformed images of the grid wire centers would increase with the increasing values of φ. If the θ values for any two consecutive grid wire centers be denoted by θ_1 and θ_2, then their spacing in the W plane is given by

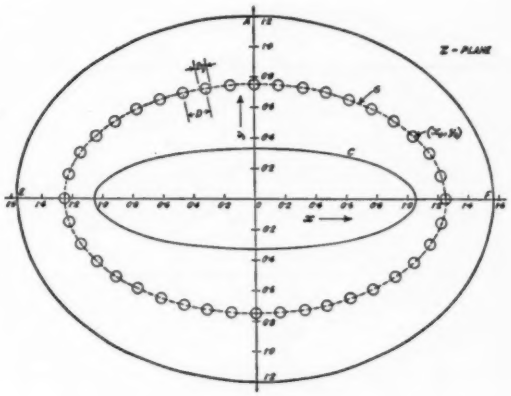


FIG. 1. Transverse section of the elliptic triode in the Z plane ($K=1$).

$$c_\varphi=\varphi_2-\varphi_1=f(\varphi)$$
$$=K(\theta_2-\theta_1)=Kf(\theta). \tag{12}$$

We now derive some approximate expressions which hold good closely for the case where the grid-wire radius (r_g) is small compared to (K/e_g)—a case usually met with in practice.

Firstly, as shown in the Appendix, for such case, a circle of radius r_g with its center lying on the ellipse G, has for its image in the W plane a figure which is very nearly a circle. This image circle has a radius given by

$$r_\varphi=r_g e_g/[1-e_g^2\sin^2(\varphi/K)]^{\frac{1}{2}}, \tag{13}$$

and its center is located at the point (φ,ψ_g).

Next we note that in the W plane, ψ_g is the same for all the grid-wire centers but φ varies. Thus the radii of the transformed grid-wire circles are larger for larger φ and hence for larger values of θ in the Z plane.

In the W plane, therefore, not only the spacing c_φ between consecutive grid-wire centers but also the grid-wire radius, r_φ increases as one goes away from the ψ axis. This is illustrated in Fig. 2(a). But, nevertheless, as shown below, the screening fraction (defined by $s=2r_\varphi/c_\varphi$ remains very approximately a constant.

Consider a section $ABCD$ of the W plane [Fig. 2(a)] bounded by the lines $\varphi=\varphi'=K\theta'$ and $\varphi=\varphi''=K\theta''$ ($\theta''>\theta'$) with the transformed grid-wire center located at $\varphi=(\varphi'+\varphi'')/2=K(\theta'+\theta'')/2=K\theta$ and $\psi=\psi_g$. For this section, $c_\varphi=K(\theta''-\theta')$ and the screening fraction of the W plane triode geometry is, from Eq. (13),

$$s=2r_g e_g/[K(\theta''-\theta')(1-e_g^2\sin^2\theta)^{\frac{1}{2}}]$$
$$=\frac{2r_g e_g}{K(\theta''-\theta')[(\partial/\partial\theta)E(\delta_g,\theta)]},$$

where $E(\delta_g,\theta)$ is the incomplete elliptic integral of the second kind. Assuming that the interval $\theta''-\theta'=\Delta\theta$ is so small that $E(\delta_g,\theta)$ may be taken to vary linearly with θ in the foregoing interval, the foregoing equation

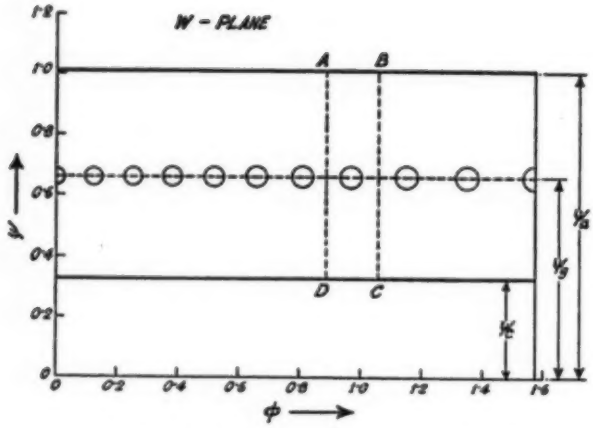


FIG. 2(a). Transverse section of the W plane representation of one quadrant of the elliptic triode ($K=1$).

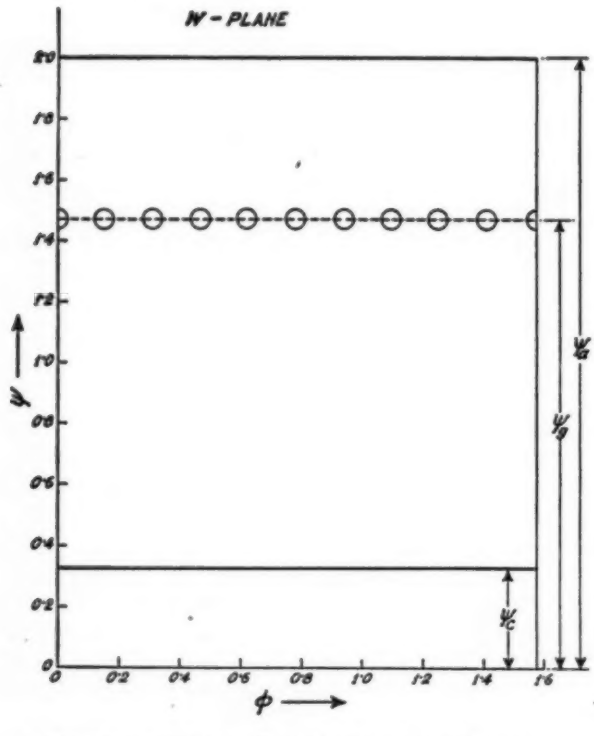


FIG. 2(b). Same as in Fig. 2(a) with small value of e_g.

reduces to

$$\begin{aligned} s &\approx 2r_g e_g/[K\Delta E(\delta_g,\theta)] \\ &= 2r_g/[(K/e_g)\{E(\delta_g,\theta'')-E(\delta_g,\theta')\}] \\ &\approx 2r_g/D \\ &= r_g e_g N/[KE], \quad \text{from Eq. (1).} \end{aligned} \tag{14}$$

It can be easily shown by actual numerical calculation that Eq. (14) holds good to a high degree of accuracy even for $e_g \doteq 1$, provided that $N \geqslant 6$. For such cases, therefore, the screening fraction of the W plane triode is constant and is given by Eq. (14).

Finally, from Eqs. (13) and (14) it follows that the grid-wire spacing should vary with φ as follows:

$$c_\varphi \approx c_0/[1-e_g^2 \sin^2(\varphi/K)]^{\frac{1}{2}}, \tag{15}$$

where $c_0 = 2KE/N = e_g D$.

It may be easily shown that for $e_g < 0.45$ the following relations hold valid within 10 percent:

$$c = K\pi/N, \quad r = r_g e_g, \quad s = 2Nr_g e_g/(K\pi).$$

Under this condition the transformed tube geometry in the W plane reduces to that of an ideal plane triode as shown in Fig. 2(b).

4. INTERELECTRODE CAPACITANCES AND AMPLIFICATION FACTOR

As pointed out in Sec. 3, the W plane image of the elliptic triode [Fig. 2(a)] can be looked upon as a plane triode of constant screening fraction but of variable pitch. No direct analytical solution for such a triode is known. It may, however, be noted that the W plane triode of Fig. 2(a) may be divided into $N/2$ symmetrical sections each containing a single grid wire. Each one of the sections, say $ABCD$, may be taken as a part of an ordinary plane triode whose screening fraction and pitch are the same as those of the section chosen. One can, therefore, apply the known results for a plane triode individually to these $N/2$ sections and obtain expressions for their interelectrode capacitances and amplification factors. The total interelectrode capacitances of the triode would then be given by the sum of these individual capacities and these, in their turn, will give the mean amplification factor of the valve.

Let us, therefore, consider one of the $N/2$ sections mentioned previously ($ABCD$) of unit height in a direction perpendicular to the plane of the figure and let the grid-wire center for this section be (φ,ψ_g). This section represents a system of electrostatic capacities which constitute a Δ network as shown in Fig. 3(a). This, again, is equivalent to the star network of Fig. 3(b). Here the common terminal represents a hypothetical equipotential plane whose electrostatic field near the cathode and the anode is the same as that of the actual grid and is called the equivalent grid sheet. Provided that $s<1/6$, $d_{cg}>c_\varphi$, $d_{ga}>c_\varphi$ elements of this star network are obtained from the theory of plane triodes as[1,4]

$$C_{1\varphi} = \frac{c_\varphi}{4\pi d_{ga} - 2c_\varphi \ln\cosh(2\pi r_\varphi/c_\varphi)} \tag{16}$$

$$C_{2\varphi} = \frac{c_\varphi}{4\pi d_{cg} - 2c_\varphi \ln\cosh(2\pi r_\varphi/c_\varphi)} \tag{17}$$

$$C_{3\varphi} = \frac{1}{2\ln\coth(2\pi r_\varphi/c_\varphi)}. \tag{18}$$

[4] W. G. Dow, *Fundamentals of Engineering Electronics* (John Wiley and Sons, Inc., New York, 1952).

Using Eqs. (9), (10), (14), and (15) and referring to the Z plane

$$C_{1\theta}=\frac{E}{[2\pi N(1-e_g^2\sin^2\theta)^{\frac{1}{2}}\ln\{\cot(\delta_a/2)/\cot(\delta_g/2)\}-2E\ln\cosh\{\pi r_g e_g N/(KE)\}]} \quad (19)$$

$$C_{2\theta}=\frac{E}{[2\pi N(1-e_g^2\sin^2\theta)^{\frac{1}{2}}\ln\{\cot(\delta_g/2)/\cot(\delta_c/2)\}-2E\ln\cosh\{\pi r_g e_g N/(KE)\}]} \quad (20)$$

$$C_{3\theta}=\frac{1}{2\ln\coth\{\pi r_g e_g N/(KE)\}}. \quad (21)$$

The capacitances C_1 and C_3 determine the relative effectiveness of the anode and the actual grid in fixing the potential of the equivalent grid sheet and hence the off-cathode field. Amplification factor of the valve is, therefore, defined as

$$\mu_\theta=C_{3\theta}/C_{1\theta}=[1/\ln\coth\{\pi r_g e_g N/(KE)\}]\times[(\pi N/E)(1-e_g^2\sin^2\theta)^{\frac{1}{2}}\ln\{\cot(\delta_a/2)/\cot(\delta_g/2)\}-\ln\cosh\{\pi r_g e_g N/(KE)\}]. \quad (22)$$

In valves as used in practice the second term in the numerator of Eq. (22) is always less than one-tenth of the value of the first. This term may, therefore, be omitted. We thus get

$$\mu_\theta/\mu_0=(1-e_g^2\sin^2\theta)^{\frac{1}{2}}, \quad (23)$$

where

$$\mu_0=(\pi N/E)\frac{\ln\{\cot(\delta_a/2)/\cot(\delta_g/2)\}}{\ln\coth\{\pi r_g e_g N/(KE)\}}, \quad (24)$$

is the amplification factor for the central region of the valve. The r.h.s. of Eq. (23) can be expressed in terms of Jacobi's elliptic function *dnu*, which is readily available in graphical and tabular form.[5] Here the variable u is the incomplete elliptic integral of the first kind $F(\delta,\theta)$. For practical handling, however, it is desirable to have a plot of the ratio (μ_θ/μ_0) directly against θ. Such a set of curves is given in Fig. 4(a). In Fig. 4(b) is given a graphical representation of Eq. (24).

In order to compute the total interelectrode capacitances of the valve, we note that the second term in the denominator of Eqs. (16) and (17) does not exceed one-tenth of the value of the first. Neglecting this term and summing up for all the $2N$ sections, we get from Eq. (16),

$$C_1=\sum\frac{c_\varphi}{4\pi K\ln\{\cot(\delta_a/2)/\cot(\delta_g/2)\}}.$$

Use of Eq. (13) gives $\sum c_\varphi=2K\pi$. Thus

$$C_1=1/[2\ln\{\cot(\delta_a/2)/\cot(\delta_g/2)\}]. \quad (25)$$

Similarly, from Eqs. (17) and (18)

$$C_2=1/[2\ln\{\cot(\delta_g/2)/\cot(\delta_c/2)\}] \quad (26)$$

$$C_3=N/\ln\coth\{\pi r_g e_g N/(KE)\}. \quad (27)$$

The average value of the amplification factor is therefore given by

$$\mu=C_3/C_1=\frac{2N\ln\{\cot(\delta_a/2)/\cot(\delta_g/2)\}}{\ln\coth\{\pi r_g e_g N/(KE)\}}. \quad (28)$$

It is interesting to note from Eqs. (24) and (28) that μ_0 approaches μ as δ_g becomes small.

From Eqs. (25) to (27) one obtains expressions for the total Δ-network capacitances:

$$C_{ac}=\frac{C_2}{\mu(1+1/\mu+1/\mu')},\qquad C_{gc}=\frac{C_2}{1+1/\mu+1/\mu'},\qquad C_{ga}=\frac{C_1}{1+1/\mu+1/\mu'}, \quad (29)$$

where

$$\mu'=C_3/C_2=\frac{2N\ln\{\cot(\delta_g/2)/\cot(\delta_c/2)\}}{\ln\coth\{\pi r_g e_g N/(KE)\}}. \quad (30)$$

5. EQUIVALENT GRID SHEET POTENTIAL AND PERVEANCE

The action of a triode valve is conveniently visualized in terms of the equivalent grid sheet potential. This potential may be obtained from the equivalent star network of the valve capacities by determining the potential of the neutral point. This potential (V_0)

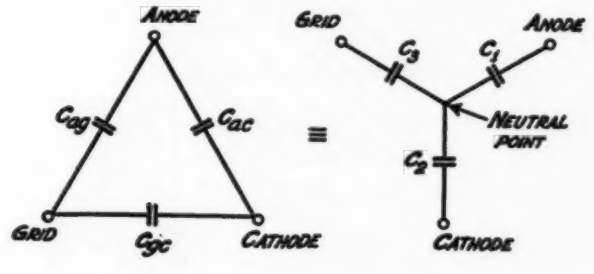


FIG. 3(a). Delta network representing the interelectrode capacitances of a symmetric section of the triode containing a single grid wire. (b) Equivalent star network of (a).

[5] E. Jahnke, and F. Emde, *Tables of Functions* (Dover Publications, New York, 1945).

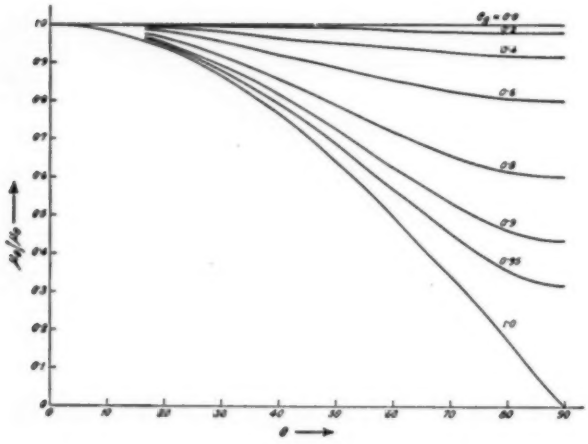


FIG. 4(a). Curve showing the variation of amplification factor with the parametric angle θ.

satisfies the condition that the net charge induced in the electrode system under purely electrostatic condition is zero, i.e.,[1,4]

$$C_1(V_a-V_0)+C_2(V_c-V_0)+C_3(V_g-V_0)=0,$$

where V_a, V_g, and V_c are the potentials of the anode, grid, and cathode, respectively.

Putting $V_c=0$, we get the equivalent grid sheet potential

$$V_0=\frac{V_g+V_a(C_1/C_3)}{1+(C_1/C_3)+(C_2/C_3)}$$

$$=\frac{V_g+V_a/\mu}{1+(1/\mu)+(1/\mu')}. \tag{31}$$

We can now apply Child's law and write

$$i=PV_0^{\frac{3}{2}}, \tag{32}$$

where i is the cathode current and P is a constant of the valve. In conformity with the nomenclature adopted for the cases of plane and cylindrical triodes, P may be called the perveance of the elliptic triode. It remains, however, to express P in terms of the geometrical constants of the valve.

A rigorous analysis of the problem is difficult as it involves the solution of a two-dimensional Poisson's equation. An approximate solution may, however, be obtained by following the method adopted by O'Neill and by Matricon and Trouve for the case of a filamentary cathode between two equidistant plane anodes.[4,6,7] The method is based on the assumption

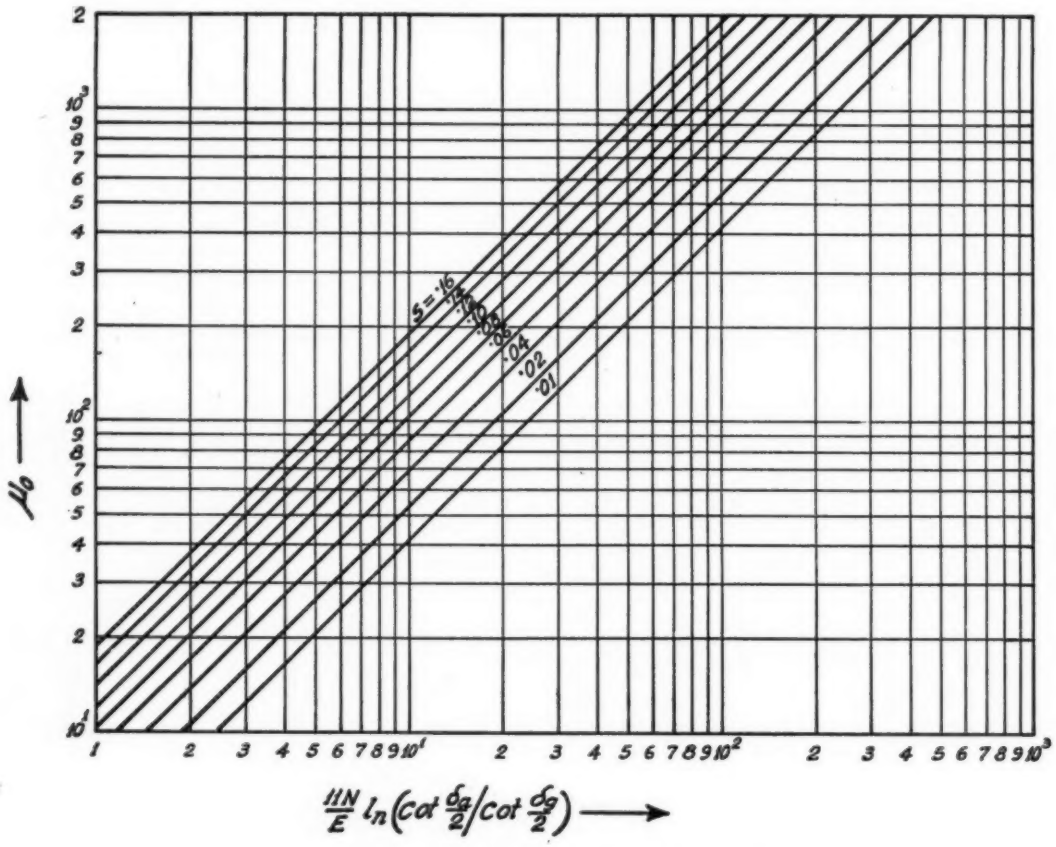


FIG. 4(b). Graphical representation of Eq. (24).

[6] G. D. O'Neill, Sylvania Technologist **3**, 22 (1950).
[7] M. Matricon and S. Trouve, Onde. Elect. **30**, 510 (1950).

that the space charge flow equations of two diodes, having the same value of electrostatic capacitance between their electrodes, are the same.

The system composed of the cathode and the equivalent grid sheet plane may be considered to act as a diode as shown in Fig. 5. From purely electrostatic consideration this represents a plane parallel capacity of value

$$\pi K/(4\pi d_{cg})=C, \text{ say}$$

per unit length in the direction normal to the plane of the figure. Using Eq. (9)

$$C=1/[4\ln\{\cot(\delta_g/2)/\cot(\delta_c/2)\}]$$

$$=\frac{1}{2\ln\left[\frac{K\cot^2(\delta_g/2)}{K\cot^2(\delta_c/2)}\right]}.$$

Thus, the capacity system composed of the semielliptic cathode and grid is equivalent to a cylindrical capacity having for its inner and outer radii

$$\left.\begin{aligned}R_i&=K\cot^2(\delta_c/2)\\R_0&=K\cot^2(\delta_g/2)\end{aligned}\right\}. \tag{33}$$

The perveance of this cylindrical diode may, therefore, be taken as the perveance of the actual triode. Remembering that the new cylindrical diode, as obtained above, represents only the upper half of the actual triode, as shown in Fig. 1, we obtain from the well-known equation of space charge flow in a cylindrical diode the perveance[1,4]

$$P=\frac{29.32\times10^{-6}\times l}{\beta^2 R_0}$$

$$=\frac{29.32\times10^{-6}\times l}{K\beta^2\cot^2(\delta_g/2)}, \tag{34}$$

where

$$\beta^2=2\ln\{\cot(\delta_g/2)/\cot(\delta_c/2)\}$$
$$-(8/5)[\ln\{\cot(\delta_g/2)/\cot(\delta_c/2)\}]^2$$
$$+(11/15)[\ln\{\cot(\delta_g/2)/\cot(\delta_c/2)\}]^3+\text{etc.}, \tag{35}$$

and l is the length of the anode.

An expression for the transconductance g_m of the elliptic triode may now be obtained by evaluating $(\partial i/\partial V_g)v_a$ from Eq. (32) and making use of Eqs. (28), (30), (34), and (35).

6. DISCUSSION

The expression for amplification factor of an elliptic triode shows that, in general, μ would depend upon the value of θ and hence would vary from section to section. The extent of this variation is determined mainly by the eccentricity of the grid ellipse. Even for $e_g=0.45$,

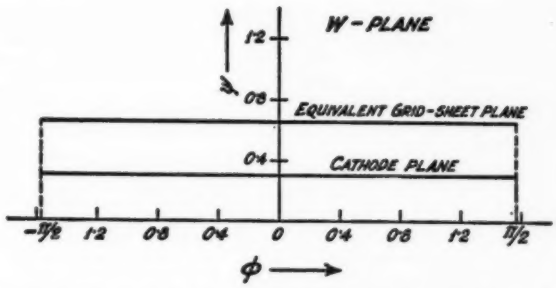


FIG. 5. Equivalent diode representation in the W plane of the upper half of the elliptic triode.

a 10 percent variation in μ will occur. As is well known, however, the value of μ as realized under actual operating condition of a valve may differ from the theoretical value by as much as ±20 percent because of the unavoidable simplifying assumptions in course of the analysis. As such, it is reasonable to take the value of μ as given by Eq. (28) as the average value of μ, at least up to $e_g=0.85$. It may also be noted that according to Eqs. (24) and (28) the values of both amplification factors, i.e., μ_0 and the average μ increase with

(i) the increase of the number of grid wires $(2N)$,
(ii) the decrease in the eccentricity of the anode ellipse (e_a),
(iii) the increase in the eccentricity of the grid ellipse (e_g),
(iv) the increase in the diameter of the grid wires (r_g).

Of these, the roles of N and e_g appear to be more critical. We further note that the value of μ is independent of the eccentricity of the cathode ellipse. This is because of the assumption that the pitch of the grid-wire is small compared to the distance between the cathode and the grid surface.

The variation of μ with θ can be minimized by making the screening fraction (s) variable. This may be achieved by varying either the grid-wire radius or the grid pitch.

If the grid-wire radius is varied then one finds from Eqs. (13), (15), (16), and (18) that for μ to remain constant for all values of θ, the following condition should be satisfied:

$$(1-e_g^2\sin^2\theta)^{\frac{1}{2}}/\ln\coth(2\pi r_\theta/D)$$
$$=(\mu E/\pi N)/\ln\{\cot(\delta_a/2)/\cot(\delta_g/2)\}=-1/B, \text{ say},$$

where r_θ is the grid-wire radius at θ. This readily gives

$$r_\theta=(D/4\pi)\ln\coth[(1/2)Bdnu], \tag{36}$$

where values of dnu are obtained from Fig. 4(a) and those of ln coth are obtainable from Smithsonian tables.[8]

If, however, the grid pitch is varied to make μ constant, then the variation of grid pitch with θ,

[8] G. F. Becker and C. E. Van Orstrand, *Smithsonian Matematic Tables, Hyperbolic Functions* (The Smithsonian Institution, 1949).

according to Eqs. (13), (15), (16), and (18), is given by the transcendental relation

$$\frac{D_\theta \ln \coth(2\pi r_g/D_\theta)}{(1-e_g^2\sin^2\theta)^{\frac{1}{2}}} = [2\pi K/(\mu e_g)] \ln\{\cot(\delta_a/2)/\cot(\delta_g/2)\} = A, \text{ say,}$$

or,

$$[\ln \coth f'(\theta)]/f'(\theta) = f(\theta), \tag{37}$$

where D_θ is the grid pitch at θ, $f(\theta)=\{A/(2\pi r_g)\}dnu$ and $f'(\theta)=2\pi r_g/D_\theta$. The solution of the above equation is given graphically in Fig. 6.

Coming to the expression for perveance one notices that it varies inversely as the focal distance K which otherwise plays the role of a mere scaling factor. Further, it varies inversely as the square of $\cot(\delta_g/2)$ which points to the crucial role played by the eccentricity of the grid ellipse in determining the performance of the valve. However, more critical discussion of perveance will be a little speculative at this stage, since it is to be remembered that all deductions regarding the same are based on a number of very tentative assumptions (originally introduced by O'Neill and by Matricon and Trouve).

In order to illustrate the practical applications of some of the deductions made let us consider the design of a 6C5-GT/G triode. This has the following electrode dimensions:

Cathode—cylindrical;	diameter	=0.15 cm
Grid—elliptic;	focal distance	=0.225 cm
	eccentricity	=0.9
	grid wire diameter	=0.0075 cm
	grid pitch	=0.06 cm
Anode—combination of a cylinder (diameter=2.0 cm) with an elliptic grid (focal distance=0.225 cm, eccentricity =0.6). Length=1.7 cm.		

With the above dimensions the following constants are obtained under typical working conditions:

$$\mu=20, \quad P=9\times10^{-4} \text{ amp (volt)}^{\frac{3}{2}}.$$

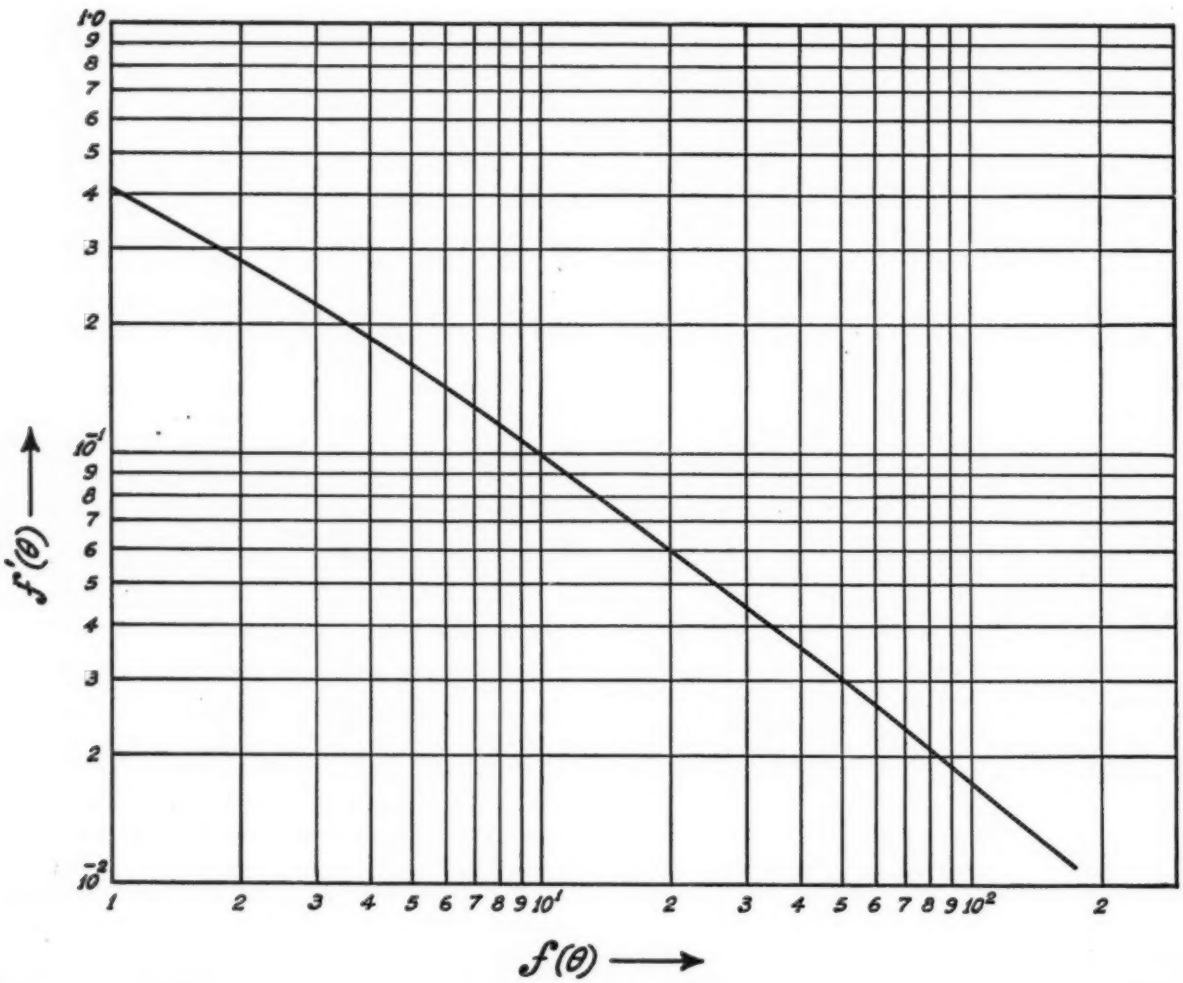


FIG. 6. Graphical solution of Eq. (37).

To realize the above valve constants from a triode whose electrodes belong to a system of confocal ellipses, we note that since we have only two equations [Eqs. (28) and (34)] at our disposal, the design of the electrode structure will not be unique. If, however, the grid structure and the active length of the elliptic triode are assumed to be the same as that of 6C5-GT/G, the dimensions of the cathode and anode, as determined from Eqs. (28) and (34), are found to be

Cathode—elliptic;	focal distance	$=0.225$ cm
	eccentricity	$=0.96$
	major axis	$=0.47$ cm
	minor axis	$=0.13$ cm
Anode—elliptic;	focal distance	$=0.225$ cm
	eccentricity	$=0.45$
	major axis	$=1.0$ cm
	minor axis	$=0.9$ cm.

The dimensions of the elliptic cathode as found in the foregoing is reasonable, for the diameter of the actual cathode which is cylindrical is found to be intermediate between the major and the minor axis of the ellipse. Further, the eccentricity 0.45 of the anode ellipse is also according to expectation, because, the combined effect of the elliptic grid and the cylinder as used in the practical valve would be to give rise to an effective anode which is approximately an ellipse occupying an intermediate position between the two.

APPENDIX. TRANSFORMATION OF THE GRID-WIRE CIRCLES FROM THE Z TO W PLANE

Let us consider a circle of radius r_0 with its center located at the point (x_0,y_0) on an ellipse in the Z plane (Fig. 7). The locus of this circle, after transformation through the use of Eq. (1), in the W plane can be derived by substituting for x and y from the relations (4a) and (4b) in the equation

$$(x-x_0)^2+(y-y_0)^2=r_0^2. \tag{i}$$

Thus

$$\begin{aligned}&\cosh^2(\psi_0/K)\sin^2(\varphi_0/K)+\sinh^2(\psi_0/K)\cos^2(\varphi_0/K)\\&+\cosh^2(\psi/K)\sin^2(\varphi/K)+\sinh^2(\psi/K)\cos^2(\varphi/K)\\&-2[\cosh(\psi/K)\cosh(\psi_0/K)\sin(\varphi/K)\sin(\varphi_0/K)\\&+\sinh(\psi/K)\sinh(\psi_0/K)\cos(\varphi/K)\cos(\varphi_0/K)\\&=(r_0/K)^2,\end{aligned} \tag{ii}$$

where (φ,ψ) is any point on the locus in the W plane, and (φ_0,ψ_0) is a point in the same plane corresponding to the center of the circle in the Z plane.

Equation (ii) may be reduced to

$$\begin{aligned}&[\cosh(\langle\psi-\psi_0\rangle/K)-\cos(\langle\varphi-\varphi_0\rangle/K)]\\&\times[\cosh(\langle\psi+\psi_0\rangle/K)+\cos(\langle\varphi+\varphi_0\rangle/K)]\\&=(r_0/K)^2.\end{aligned} \tag{iii}$$

The locus represented by Eq. (iii) is a closed curve and approximates a circle when r_0 is small. Thus putting

$$\psi+\psi_0\approx 2\psi_0 \quad \text{and} \quad \varphi+\varphi_0\approx 2\varphi_0,$$

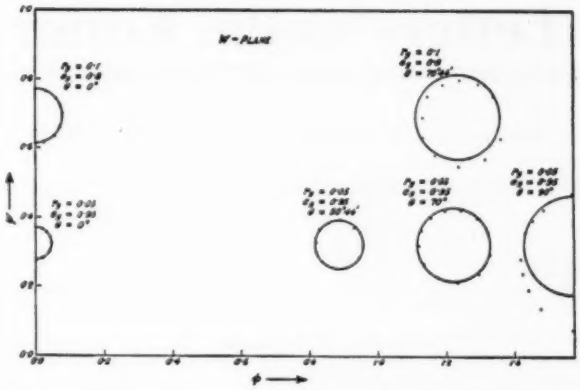


FIG. 7. W plane images of grid-wire circles ($K=1$). Solid line curves from Eq. (iv). Points obtained from Eq. (vi) are also shown.

and expanding $\cosh(\langle\psi-\psi_0\rangle/K)$ and $\cos(\langle\varphi-\varphi_0\rangle/K)$ by Maclaurin's theorem and neglecting higher-order terms, the equation of the locus in the W plane becomes

$$(\psi-\psi_0)^2+(\varphi-\varphi_0)^2=\left[\frac{r_0}{[\cosh^2(\psi_0/K)-\sin^2(\varphi_0/K)]^{\frac{1}{2}}}\right]^2.$$

Substituting for $\psi_0=K\cosh^{-1}(1/e_0)$,

$$(\psi-\psi_0)^2+(\varphi-\varphi_0)^2=\frac{r_0e_0}{[1-e_0^2\sin^2(\varphi_0/K)]^{\frac{1}{2}}}. \tag{iv}$$

Equation (iv) represents a circle in the W plane with its center located at (φ_0,ψ_0) and having a radius given by

$$r=r_0e_0/[1-e_0^2\sin^2(\varphi_0/K)]^{\frac{1}{2}}. \tag{v}$$

To test the range of validity of Eq. (iv) a point-by-point transformation of the circle given by (i) was carried out from the Z to W plane. The required relations are given explicitly by

$$\left.\begin{aligned}\psi&=K\cosh^{-1}\{(1/\langle 2K\rangle)[(x+K)^2+y^2]^{\frac{1}{2}}\\&\qquad+(1/\langle 2K\rangle)[(x-K)^2+y^2]^{\frac{1}{2}}\}\\\varphi&=K\sin^{-1}\{(1/\langle 2K\rangle)[(x+K)^2+y^2]^{\frac{1}{2}}\\&\qquad-(1/\langle 2K\rangle)[(x-K)^2+y^2]^{\frac{1}{2}}\}\end{aligned}\right\}. \tag{vi}$$

The resultant loci obtained by such a procedure together with those obtained from Eq. (iv) are shown in Fig. 7 for several values of e_0 and r_0. It is seen that the agreement between the two is quite satisfactory, provided that r_0 is small in comparison to K/e_0.

ACKNOWLEDGMENT

The present work was carried out in the Institute of Radio Physics and Electronics, Calcutta University. The authors wish to thank Professor S. K. Mitra for his encouragement, helpful criticism, and guidance in course of preparation of the paper.

Letters to the Editor

Formation of Single Crystal Silicon Fibers

EVERETT R. JOHNSON AND JAMES A. AMICK
David Sarnoff Research Center, RCA Laboratories Division, Princeton, New Jersey
(Received April 27, 1954)

THE formation of metal fibers or "whiskers" is an unusual occurrence, especially when the formation of these fibers involves a complicated reaction in the gas phase. Normally, when $SiCl_4$ is reacted with zinc vapor at high temperature, crystals of silicon of macroscopic size are obtained.[1] However, we have observed that if the $SiCl_4$ is diluted with a carrier gas, the resulting silicon consists principally of fibers.

Experimental.—The fibers were prepared by saturating a carrier gas (argon or hydrogen) with $SiCl_4$ at room temperature and reacting this mixture with Zn vapor at about 800°–1000°C (Fig. 1). The fibers form along the walls and at the exit end of the tube (indicated by areas marked *A* and *B*, Fig. 1). All chemicals were C. P. grade and used without further purification.

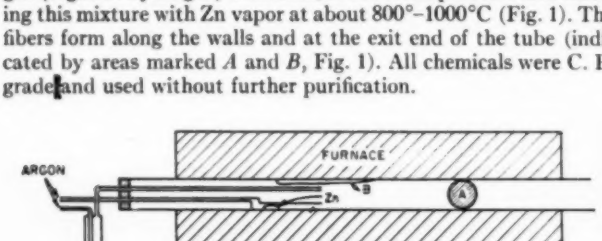


FIG. 1. Furnace used in the preparation of silicon fibers.

The fibers were examined with the experimental high voltage electron microscope, which has been previously described.[2] This microscope permits the placing of the specimen directly below the projector lens and in this position diffraction and shadow micrographs can be obtained. Since there are no magnetic lenses between the specimen and the viewing screen or plate, the orientation of the diffraction pattern (if any) is given by rectilinear projection from the specimen.

Discussion.—The silicon fibers, when first formed, are brown in color and are heavily contaminated with Zn, $ZnCl_2$, and some ZnO. The ZnO is believed to be formed by back diffusion of air during the reaction. The Zn and the $ZnCl_2$ are easily removed from the fibrous material by heating in a vacuum to about 1000°C.

Figure 2 shows the shadow micrograph of a tuft of fibers which were separated from a large mass of material at random, and Fig. 3 shows the corresponding diffraction pattern. The absence of

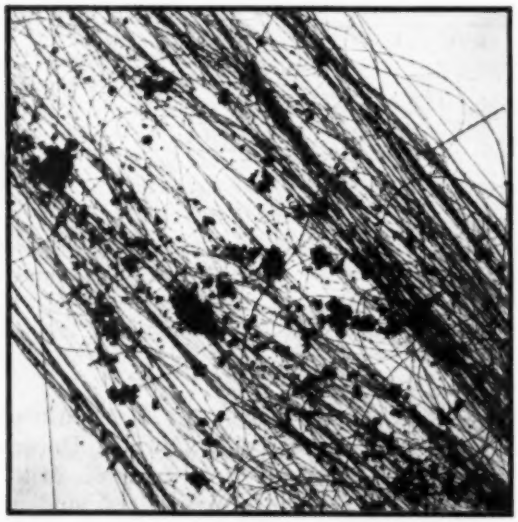


FIG. 2. Oriented silicon metal fibers—magnification about 20×.

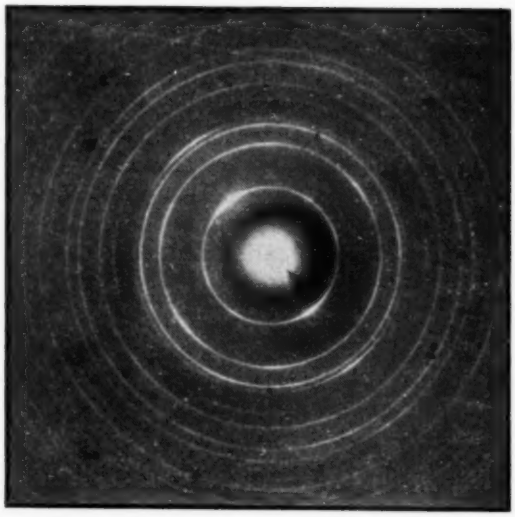


FIG. 3. Diffraction pattern of a portion of the area seen in Fig. 2.

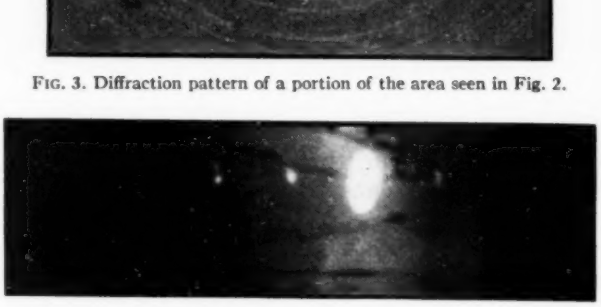


FIG. 4. Diffraction pattern of single fiber of silicon.
(The fiber running through the central spot gives rise to the pattern.)

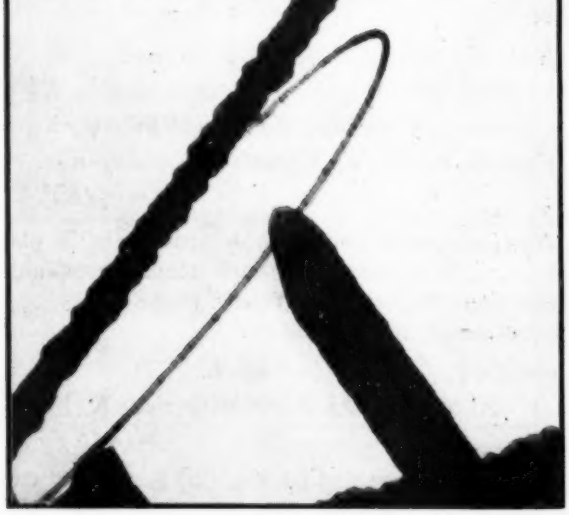


FIG. 5. Electron micrographs of single fibers.
(Magnification about 4800×.)

TABLE I.

hkl	Electron diffraction d in A.	Silicon patterns (x-ray diffraction) from ASTM card index 3-0549	3-0544	From NBS circular 539 NBS
110		3.81		
111	3.10	3.12	3.11	3.138
200	2.70			
220	1.90	1.90	1.91	1.920
311	1.61	1.63	1.63	1.638
222	1.55	1.56		
400	1.34	1.35	1.35	1.357
331	1.23	1.24	1.24	1.246
420	1.20			
422	1.09	1.10	1.10	1.1083
333 511	1.03	1.04	1.04	1.0450
440	0.95	0.958	0.958	0.9599
531	0.91	0.915	0.916	0.9178
443 540 621	0.85	0.856	0.857	0.8586
533	0.82	0.828	0.828	0.8281

diffuse rings in the diffraction pattern indicates that no amorphous SiO_2 is present. The orientation of the pattern indicates that its origin is the fibrous material. Furthermore, the intensity of the pattern along the direction of the fibers indicates strongly that these fibers are single crystals. This is substantiated by patterns obtained on individual fibers (Fig. 4). The light spots lie on a line in the direction of the long fiber axis and the intervals correspond to the distance between 111 planes in silicon. The fibers thus grow along the 111 axis.

The d values calculated from the diffraction pattern of Fig. 3 are listed in Table I, and compared with those given for silicon in the ASTM index and the National Bureau of Standards report. These latter were obtained with x-radiation, for which the reflections from the 222, 200, and 420 planes are forbidden.

The random coating on the specimen in Fig. 2 does not give a separate diffraction pattern and may possibly be silicon. Zinc oxide has been identified in diffraction patterns from other preparations and zinc, either free or chemically combined, has been established as present by x-ray fluorescence in the bulk material from which the specimen in Fig. 3 was taken. It is also possible that the random coating in Fig. 2 is zinc oxide, but its concentration in the beam was insufficient to give rise to a separate diffraction in this case.

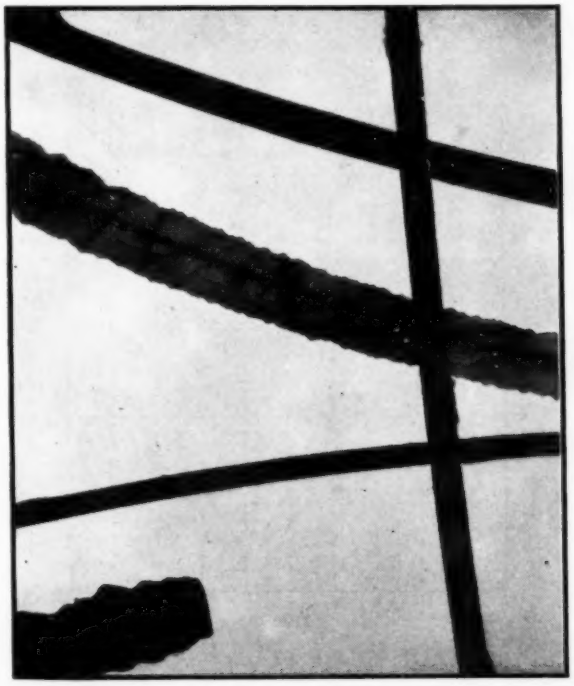


FIG. 6. Showing roughness of the edges and interference phenomena on single fibers. (Magnification about 8000×.)

Figures 5 and 6 are electron micrographs of individual fibers; the lines appearing in the fibers are presumed to be interference phenomena.

The fibers have diameters averaging about 1 micron and lengths up to 1 cm. Most of the fibers appear to have rough edges under sufficient magnification and the internal structure of some fibers indicates that spiral dislocations may play a role in their growth.

Diffraction patterns taken on the material within a few days of its preparation showed no visible changes; after five months, however, no pattern could be obtained, indicating that the surface of the silicon had changed at least to the depth of penetration of the electron beam.

The authors wish to acknowledge stimulating discussions with Dr. James Hillier.

[1] Lycer, Alser, and Lewis, Trans. Electrochem. Soc., 359–363 (1949).
[2] Congres de Microscopie Electronique Editions de la Revue d'Optique, Paris, 1952.

Laminar Flow of an Electrically Conducting Incompressible Fluid in a Circular Pipe*

SHIH I. PAI
Institute for Fluid Dynamics and Applied Mathematics, University of Maryland, College Park, Maryland
(Received February 8, 1954)

RECENTLY magnethydrodynamics, the flow problems of electrically conducting fluid where electromagnetic forces are of the same order of magnitude as pressure forces and viscous forces, has attracted great attention. Magnetohydrodynamics is of importance in problems of astrophysics, geophysics, and the behavior of interstellar gas masses. It is interesting to astrophysicists as well as aerodynamists.[1]

In studying the magnetohydrodynamics, one has to deal with the hydrodynamical equations and the electromagnetic equations simultaneously. It is well known that there is no general method of solution for hydrodynamical equations alone. Thus it is much more difficult to investigate magnetohydrodynamics. One of the first steps to understand the magnetohydrodynamics is to find some simple exact solutions of magnetohydrodynamical equations in order to bring out the essential feature of magnetohydrodynamics. In this note the exact solution of the steady laminar flow of an electrically conducting, viscous and incompressible fluid in a circular pipe under an external radial magnetic field is given.

The fundamental equations of a magnetohydrodynamics of an incompressible fluid are

$$\frac{\partial \mathbf{q}}{\partial t}+(\mathbf{q}\cdot\nabla)\mathbf{q}=-\frac{1}{\rho}\nabla p+\frac{\rho_e}{\rho}\mathbf{E}+\frac{\mu_e}{\rho}\mathbf{J}\times\mathbf{H}+\Delta\nabla^2\mathbf{q}, \tag{1}$$

$$\nabla\cdot\mathbf{q}=0, \tag{2}$$

$$\nabla\times\mathbf{H}=\mathbf{J}+(\partial\epsilon\mathbf{E}/\partial t), \tag{3}$$

$$\nabla\times\mathbf{E}=-(\partial\mu_e H/\partial t), \tag{4}$$

$$\mathbf{J}=\sigma(\mathbf{E}+\mu_e\mathbf{q}\times\mathbf{H})+\rho_e\mathbf{q}, \tag{5}$$

$$\nabla\cdot\mathbf{J}+(\partial\rho_e/\partial t)=0, \tag{6}$$

where $\mathbf{q}$ is the velocity vector, p is the pressure, ρ is the density of the fluid which is constant here, ρ_e is the excess electric charge, $\mathbf{E}$ is the electrical field strength, μ_e is the magnetic permeability, $\mathbf{J}$ is the electric current density, $\mathbf{H}$ is the magnetic field strength, Δ is the coefficient of kinematic viscosity, ϵ is the dielectric constant, and σ is the electric conductivity. The MSK united system is used. ∇ is the gradient operator, $\nabla\cdot\nabla=\nabla^2$, and t is the time. Gibbs vector notations are used. Equations (1) and (2) are the hydrodynamical equations, while the rest, (3) to (6), are the electromagnetic equations. There are coupling terms in these equations such as $(\mu_e/\rho)\mathbf{J}\times\mathbf{H}$ in Eq. (1) and $\sigma\mu_e\mathbf{q}\times\mathbf{H}$ in Eq. (5), etc.

In the general three-dimensional flow, each vector has three components. Therefore, there are 14 unknowns in the present

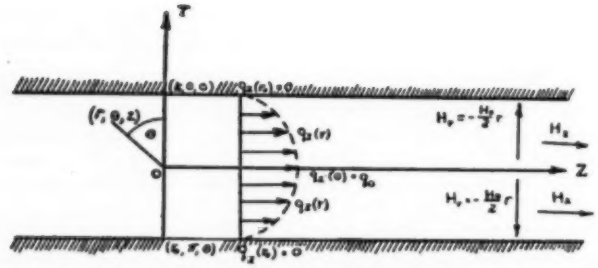


FIG. 1. Laminar flow of an electrical conducting fluid in a circular pipe under external radial magnetic field H_0.

problem, i.e., $\mathbf{q}$, $\mathbf{E}$, $\mathbf{H}$, $\mathbf{J}$, ρ_e, and p, which should be obtained from the 14 relations of Eqs. (1) to (6) under given initial and boundary conditions. The general solution is indeed very difficult.

Now we consider the simple case of laminar steady flow of an electrically conducting incompressible fluid in a nonconducting circular pipe under an external radial magnetic field which is independent of the axial position of the pipe. We assume that there is no excess electric charge ($\rho_e=0$), that the pipe is infinitely long, and that both the velocity and the magnetic fields are axially symmetrical, i.e., independent of the angular position and without tangential components. Furthermore, we assume that the velocity has only an axial component which depends entirely on the radial distance. We use the cylindrical coordinates r, θ, z (Fig. 1), where $z=0$ is arbitrarily chosen. In the present problem we have the following relations

$$\begin{gathered} q_r=q_\theta=0, \quad q_z=q_z(r) \\ H_\theta=0, \quad H_r=H_r(r), \quad H_z=H_z(r,z), \quad p=p(r,z) \\ \partial(\ )/\partial t=\partial(\ )/\partial\theta=0, \end{gathered} \tag{7}$$

where subscripts, r, θ, z refer to components along r, θ, z directions, respectively, $f(r)$ means function of r only, and $f(r,z)$ means function of both r and z. ( ) means any quantity of the unknowns.

Substituting (7) into (1) to (6), eliminating the other unknowns excpet those given in (7), and simplifying, we find the following four relations for the unknowns q_z, p, H_r, and H_z in cylindrical coordinates[2]:

$$\frac{\partial H_r}{\partial r}+\frac{H_r}{r}+\frac{\partial H_z}{\partial z}=0 \tag{8}$$

$$\frac{1}{\rho}\frac{\partial p}{\partial r}=-\frac{\mu_e}{\rho}H_z\frac{\partial H_z}{\partial r} \tag{9}$$

$$\frac{d^2q_z}{dr^2}+\frac{1}{r}\frac{dq_z}{dr}=\frac{1}{\rho\Delta}\frac{\partial p}{\partial z}-\frac{\mu_e}{\rho\Delta}H_r\frac{\partial H_z}{\partial r} \tag{10}$$

$$q_z\frac{\partial H_z}{\partial z}-H_r\frac{dq_z}{dr}=\frac{1}{\mu_e\sigma}\left(\frac{\partial^2H_z}{\partial r^2}+\frac{1}{r}\frac{\partial H_z}{\partial r}+\frac{\partial^2H_z}{\partial z^2}\right). \tag{11}$$

For instance, Eq. (8) may be derived from Eq. (4) by taking divergence of (4) and assuming that at certain instant $\nabla\cdot H=0$. From Eq. (8) we see that the simple solution which satisfies our boundary conditions that on the surface of the pipe the value of H_r is a constant and that magnetic field is finite inside the pipe including the axis of the pipe is as follows:

$$H_r=-\frac{H_0}{2}r, \quad H_z=H_0z+h(r), \tag{12}$$

where H_0 is a constant given by the value of H_r on the surface of the pipe. $h(r)$ is that part of H_z which is to be determined. It is evident that (12) satisfies (8), where $\partial h(r)/\partial z=0$.

Substituting (12) into (11), and integrating with respect to r we have

$$\frac{H_0\mu_e\sigma}{2}q_z=\frac{1}{r}\frac{dh}{dr}. \tag{13}$$

In obtaining (13), the condition that at $r=0$, $dh/dr=0$ is used.

Differentiating (9) with respect to z and using the relation (12), we have

$$\frac{1}{\rho}\frac{\partial^2p}{\partial r\partial z}=-\frac{\mu_e}{\rho}H_0\frac{dh}{dr}. \tag{14}$$

Differentiating (10) with respect to r, and using the relations (13) and (14), we have the differential equation for q_z as follows:

$$\frac{d^3q_z}{dr^3}+\frac{1}{r}\frac{d^2q_z}{dr^2}-\left(\frac{k^2r^2}{4}+\frac{1}{r^2}\right)\frac{dq_z}{dr}=0, \tag{15}$$

where

$$k^2=\frac{\mu_e^2H_0^2\sigma}{\rho\Delta}. \tag{16}$$

The solution of (15) is[3]

$$\frac{dq_z}{dr}=AI_{\frac{1}{2}}(Z_1), \tag{17}$$

where

$$Z_1=kr^2/4, \tag{18}$$

and $I_{\frac{1}{2}}$ is the modified Bessel function of order $\frac{1}{2}$ of first kind. The second kind of modified Bessel function of order $\frac{1}{2}$ can not be used here because of the condition that at $Z_1=0$, $dq_z/dr=0$. A is an arbitrary constant to be determined later.

By integration of (17) with respect to r and with the boundary conditions

$$r=0(Z_1=0), \quad q_z=q_0, \quad r=r_0(Z_1=Z_0), \quad q_z=0 \tag{19}$$

we have

$$\frac{q_z}{q_0}=1-\frac{\int_0^{Z_1}\frac{\sinh Z_1}{Z_1}dZ_1}{\int_0^{Z_0}\frac{\sinh Z_1}{Z_1}dZ_1}, \tag{20}$$

where q_0 is the velocity on the axis of the pipe.

When k is very small, i.e., $k\to0$, we have

$$q_z/q_0=1-(r^2/r_0^2), \tag{21}$$

which is the well-known parabolic velocity distribution of Poiseuille flow in a circular pipe.

When k is very large, i.e., $k\to\infty$, we have

$$\lim_{k\to\infty}\frac{\int_0^{Z_1}\frac{\sinh Z_1}{Z_1}dZ_1'}{\int_0^{Z_0}\frac{\sinh Z_1}{Z_1}dZ_1}\to0, \tag{22}$$

except $(r_0-r)\to0$. Hence,

$$q_z\cong q_0, \tag{23}$$

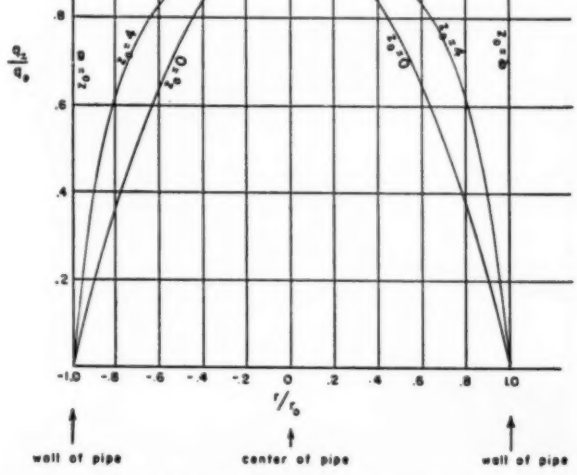


FIG. 2. Velocity distributions in a circular pipe.

except $(r_0-r)0$. For very large value of k, the velocity distribution is almost constant in the central portion of the pipe and then drops very rapidly to zero near $r=r_0$. The reason is due to the fact that the magnetic field increases the total shearing stress of the flow field. Since the shearing stress due to the magnetic field is proportional to the velocity q_z, near the wall the total shearing stress will be largely produced by the viscous force only. As a result there will be a large velocity gradient near the wall. This effect is qualitatively similar to that of turbulent flow. In turbulent flow the total shearing stress of the flow is much larger than that of the corresponding laminar flow. But the turbulent stress will be zero at the wall; thus there will be a large velocity gradient near the wall of turbulent flow to respond the large shearing stress of the flow field.[4]

Figure 2 shows three typical velocity distributions for laminar steady flow in a circular pipe under external radial magnetic field of different strengths.

After the velocity distributions $q_z(r)$ is known, H_z and p can be obtained immediately by simple quadrature. One interesting point is that the variation of pressure in the z direction is always linear no matter whether there is a magnetic field or not. Only with the external magnetic field $H_0\neq 0$, the pressure varies with the radial distance, while in ordinary laminar Poiseuille flow, the pressure is constant for a given section. This point is also similar to that of turbulent flow.[4]

The similar problem for two-dimensional case has been solved by Lehnert,[5] but there seems a misprint for the expression of velocity distribution of plane Poiseuille flow in reference 5. The correct expression for velocity distribution should be

$$\frac{q_z}{q_0}=\frac{\cosh kr_0-\cosh kr}{\cosh kr_0-1}, \tag{24}$$

where $2r_0$ is the distance between the two parallel plates and the rest symbols have similar meanings as those used in this note. The general behavior of the velocity distributions of these two problems are the same.

* This work was carried out under Contract AF18(600)-993, sponsored by the Air Research and Development Command.

[1] "Problems of cosmical aerodynamics," Proc. of the Symposium of the Motion of Gaseous Masses of Cosmical Dimensions, Paris, August 16–19, 1949. Intern. Union Theoret. Appl. Mech. and Intern. Astron. Union. Published by Central Air Documents Office, U. S. A. (1951).
[2] S. Chandrasekhar, Proc. Roy. Soc. (London) **216**, 293–309 (1953).
[3] G. N. Watson, *Theory of Bessel Functions* (Cambridge University Press, Cambridge, 1944).
[4] S. I. Pai, J. Franklin Inst. **256**, No. 4, 337–352 (1953).
[5] Bo. Lehnert, Arkiv Fysik **5**, 1–2, 69–90 (1952).

Ultimate Vacuum in a Vacuum-Enclosed Ionization Gauge

L. J. VARNERIN, JR., AND D. WHITE *
Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania
(Received May 22, 1954)

IN a typical ultra-high vacuum system[1] used in this laboratory, the lowest pressures are obtained by pumping with a Bayard-Alpert ionization gauge after the system is isolated from the pump. Recently,[1] it has been shown that the lowest obtainable pressures (5×10^{-11} to 10^{-10} mm equivalent nitrogen pressure) were limited by the rate at which the ionization gauge can pump the atmospheric helium which diffuses through the glass. If the atmospheric helium can be prevented from coming into contact with the glass, it should be possible to obtain even higher vacua.

In order to test this prediction, a Bayard-Alpert ionization gauge was enclosed within a vacuum. The experimental tube is shown schematically in Fig. 1. The tubulation of the enclosed ionization gauge is sealed to a kovar tube which, in turn, is brazed to the tubulation of an all-metal valve. The large glass jacket which encloses the gauge is sealed to the kovar tube. When the outer jacket is evacuated the only path for direct gas leakage into the gauge is through the kovar tube or the copper valve body.

The complete tube assembly was baked out several times at 475°C. for twelve-hour periods. After each bakeout the ionization gauge was outgassed by electron bombardment. The entire vacuum system was iolated from the diffusion pump and pumped by an ionization gauge to about 10^{-9} mm. At this point the enclosed gauge was isolated from the vacuum system by closing the valve shown in Fig. 1. It was then possible to pump the enclosed gauge ionically to much lower pressures. In this region of very low gas ionization currents, a background photoelectric current[1] interferes with pressure measurements. The emission electrons striking the grid cause an emission of soft x-rays, which in turn eject electrons from the ion collector. This photoelectric or soft x-ray current leaving the collector adds to the gas ionization current.

Under the most favorable circumstances encountered in our experiments, it is believed that the soft x-ray current constituted the major contribution to the observed current. Curve 1 of Fig. 2 is a plot of the observed collector current as a function of grid potential. In a standard ionization gauge employing a cylindrical collector, x-ray currents are much larger. Curves similar to curve 1 have been obtained for the standard gauge under conditions where the Bayard-Alpert ionization gauge measurements have shown gas ionization currents to be negligible.

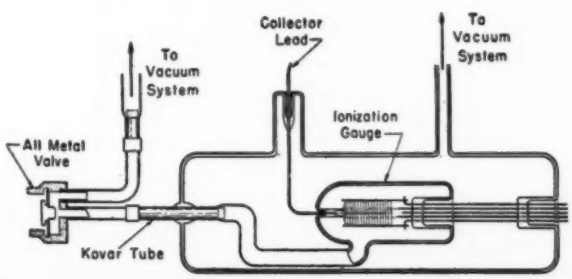


FIG. 1. Vacuum enclosed ionization gauge.

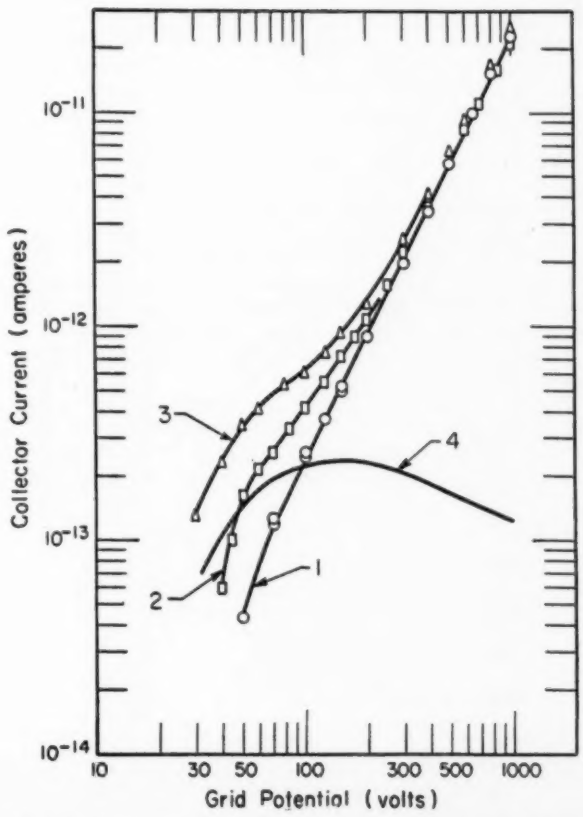


FIG. 2. Ionization gauge collector current as a function of grid voltage.

Curves 2 and 3 of Fig. 2 were obtained under less favorable conditions. Here it is clear that a current similar in nature to a typical gas ionization curve (shown as curve 4) has been added to curve 1. Curve 2 shows a current at 150-volt grid potential greater than that of curve 1 by about 4×10^{-13} amp. With an emission current of 1 ma, this corresponds to an equivalent nitrogen pressure of 4×10^{-11} mm. Curve 3 corresponds to 4×10^{-11} mm. It appears justifiable to conclude that under the conditions for which curve 1 was obtained, the total residual pressure including both adsorbable and permanent gases was substantially less than 10^{-11} mm.

It is clear from Fig. 2 that the use of lower grid voltages for the measurements of ion current gives a larger ratio of ion current to x-ray current. For measurements of pressure in the 10^{-11} mm range a definite improvement in performance of the ionization gauge is possible by reducing the grid voltage to 70 volts from the standard value of 150 volts.

These experiments have demonstrated that much lower pressures can be obtained by the positive elimination of the atmospheric helium leak. In addition, they have demonstrated that kovar and copper are vacuum tight to within the limits of measurement. They point further to the desirability of employing lower values of grid potentials to reduce the relative effect of x-ray current in an ionization gauge.

* Now at the University of Wisconsin, Madison, Wisconsin.
[1] D. Alpert, J. Appl. Phys. **24**, 860 (1953).

Note on the Propagation of Explosion-Produced Air Shocks

M. Birk, Y. Manheimer, and G. Nahmani
Scientific Department, Israeli Ministry of Defence, Hakirya, Tel-Aviv, Israel
(Received June 7, 1954)

In a recent paper, Lin[1] developed a theory for the propagation of cylindrical shock waves produced by instantaneous release of energy along an infinite straight line and concluded that the reciprocal shock velocity is a linear function of the square root of the time elapsed from the moment of energy release.

We would like to report some experimental results, which, at least within a certain range, appear to support Lin's theory.

We measured the propagation of air shocks produced by a cylindrical charge of PETN (outer radius 15 mm, inner radius 4 mm, height 30 mm) using ionization probes to signal the arrival of the shock by blanking a spiral beam of an oscillograph.[2] It was thus possible to measure the arrival of the shock at several points in the same explosion. The probes were placed at different distances from the charge in its equatorial plane, and it may therefore be assumed that, in a limited range, axial symmetry exists (this assumption has been confirmed by high-speed photographs of the explosion).

Plotting $1/U$ *versus* $t^{\frac{1}{2}}$ (Fig. 1), the shock velocity U being obtained from the distance-time curve, a fairly linear relation was obtained in the range of Mach numbers 7–14. As mentioned by Lin, the failure of this relation for higher Mach numbers is probably caused by the fact that the specific heat of air is no longer constant for this range. The deviation from linearity, noticeable from Mach number 7 downwards instead, as predicted by Lin, from Mach number 3, may well be caused by the finite dimensions of our charge.

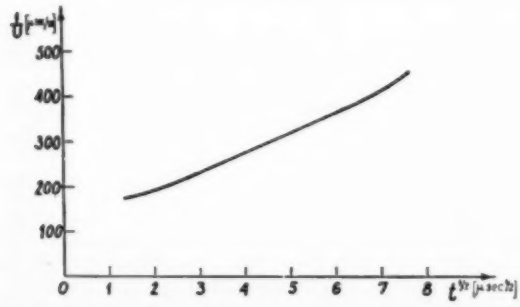


Fig. 1. Reciprocal shock velocity *versus* square root of travel time.

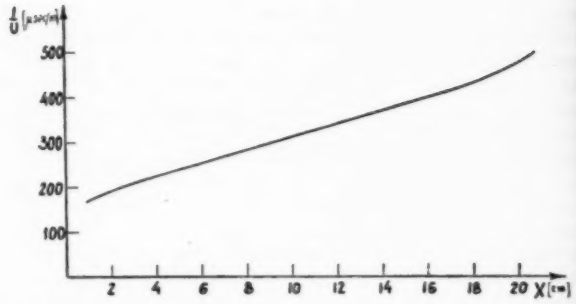


Fig. 2. Reciprocal shock velocity *versus* distance from charge surface.

On the other hand, experimental results reported by Savitt and Stresau[3] show the existence of a linear relation between $1/U$ and x (distance from charge surface) in the case of shock waves produced by highly confined cylindrical charges and by spherical charges. Our experiments on shock waves produced by unconfined cylindrical charges in air give also a substantially linear relation between these variables in the range of Mach numbers 7–14 (Fig. 2).

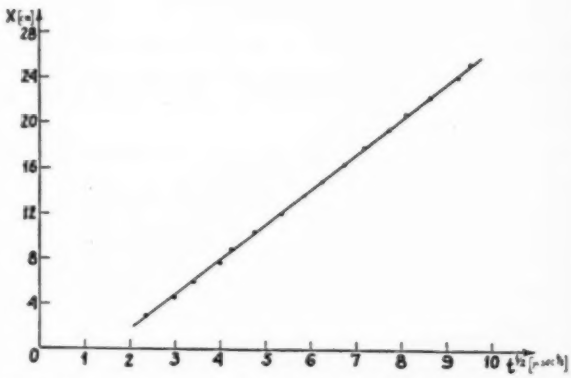


Fig. 3. Distance from charge surface *versus* square root of travel time.

It can, therefore, be expected that in the above range the distance of the shock front from the charge surface will be a linear function of $t^{\frac{1}{2}}$. This is borne out by Fig. 3 which gives experimental results obtained in the range of Mach numbers 5–14.

[1] S. C. Lin, J. Appl. Phys. **25**, 54 (1954).
[2] Birk, Erez, Manheimer, and Nahmani, Compt. rend. **238**, 654 (1954).
[3] J. Savitt and R. H. F. Stresau, J. Appl. Phys. **25**, 89 (1954).

Errata: Velocity Attenuation of Explosive-Produced Air Shocks

Jacob Savitt and R. H. F. Stresau
U. S. Naval Ordnance Laboratory, White Oak, Maryland
[J. Appl. Phys. **25**, 89–91 (1954)]

The captions under Figs. 5 and 6 of this article should read, respectively, as follows:

Instead of "Fig. 5. One-dimensional attenuation of RDX . . .." read: "Fig. 6. One-dimensional attenuation of TNT . . ."

Instead of "Fig. 6. One-dimensional attenuation of TNT . . ." read: "Fig. 5. One-dimensional attenuation of RDX . . .."